DOCUMENT RESUME

ED 219 834 CS 503 929

TITLE	Speech Research: A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for Its Investigation, and Practical Applications, Status Report, April 1-June 30, 1982.
INSTITUTION	Haskins Labs., New Haven, Conn.
SPONS AGENCY	National Institutes of Health (DHHS), Bethesda, Md.; National Inst. of Child Health and Human Development (NIH), Bethesda, Md.; National Inst. of Neurological and Communicative Disorders and Stroke (NIH), Bethesda, Md.; National Science Foundation, Washington, D.C.
REPORT NO	SR-70 (1982)
PUB DATE	82
CONTRACT	NICHHD-N01-HD-1-2420
GRANT	NICHHD-HD-01994; NIH-RR-05596; NINCDS-NS13870; NSF-PRF-8006144
NOTE	315p.
EDRS PRICE	MF01/PC13 Plus Postage.
DESCRIPTORS	*Acoustic Phonetics; *Articulation (Speech); *Communication Research; Consonants; Deafness; Elementary Education; *Motor Reactions; Perception; *Reading Difficulties; Reading Instruction; Speech Handicaps; *Speech Pathology; Speech Skills; Vowels

ABSTRACT

Research reports on the nature of speech, instrumentation for its investigation, and practical applications of research are provided in this status report covering the period of April 1 through June 30, 1982. The 13 reports deal with the following topics: (1) the functional significance of physiological tremor, (2) differences between experienced and inexperienced listeners to deaf speech, (3) a language-oriented view of reading and its disabilities, (4) phonetic factors in letter detection, (5) categorical perception, (6) short term recall by deaf signers of American Sign Language, (7) the use of phonological awareness and verbal short term memory to detect reading problems, (8) initiation versus execution time during manual and oral counting by stutterers, (9) trading relations in the perception of speech by 5-year-old children, (10) the role of the strap muscles in pitch lowering, (11) phonetic validation of distinctive features in French, (12) consonant and syllable boundaries, and (13) vowel information in postvocalic frictions. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED219834

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

SR-70 (1982)

Status Report on

SPEECH RESEARCH

A Report on the Status and Progress of Studies on the Nature of Speech, Instrumentation for its Investigation, and Practical Applications

1 April - 30 June 1982

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06510

Distribution of this document is unlimited.

(This document contains no information not freely available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.)

CS503929

ACKNOWLEDGMENTS

The research reported here was made possible in part by support from the following sources:

National Institute of Child Health and Human Development
Grant HD-01994
Grant HD-16591

National Institute of Child Health and Human Development
Contract N01-HD-1-2420

National Institutes of Health
Biomedical Research Support Grant RR-05596

National Science Foundation
Grant PRF-8006144
Grant BNS-8111470

National Institute of Neurological and Communicative Disorders and Stroke
Grant NS13870
Grant NS13617
Grant NS18010

HASKINS LABORATORIES

Personnel in Speech Research

Alvin M. Liberman,* President and Research Director
Franklin S. Cooper,* Associate Research Director
Patrick W. Nye, Associate Research Director
Raymond C. Huey,* Treasurer
Alice Dadourian, Secretary

Investigators

Arthur S. Abramson*
Peter J. Alfonso*
Cinzia Avesani[1]
Thomas Baer
Alan Bell[2]
Fredericka Bell-Berti*
Catherine Best*
Gloria J. Borden*
Susan Brady*
Robert Crowder*
Carol A. Fowler*
Louis Goldstein*
Vicki L. Hanson
Katherine S. Harris*
Alice Healy*
Kiyoshi Honda[3]
Leonard Katz*
J. A. Scott Kelso
Andrea G. Levitt*
Isabelle Y. Liberman*
Leigh Lisker*
Virginia Mann*
Charles Marshall
Ignatius G. Mattingly*
Nancy S. McGarr*
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Elliot Saltzman
Donald P. Shankweiler*
Michael Studdert-Kennedy*
Betty Tuller*
Michael T. Turvey*
Robert Verbrugge*

Technical and Support Staff

Eric L. Andreasson
Margo Carter
Elizabeth P. Clark
Vincent Gulisano
Donald Hailey
Sabina D. Koroluk
Bruce Martin
Nancy O'Brien
Marilyn K. Parnell
William P. Scully
Richard S. Sharkany
Leonard Szubowicz
Edward R. Wiley
Mary-Anne Wolf*
David Zeichner

Students*

Suzanne Boyce
André Cooper
Tova Clayman
Steven Eady
Jo Estill
Laurie B. Feldman
Nancy Fishbein
Carole E. Gelfer
Janette Henderson
Charles Hoequist
Robert Katz
Rena Krakow
Peter Kugler
Gerald Lame
Anthony Levas
Harriet Magen
Sharon Manuel
Richard McGowan
Suzi Pollock
Brad Rakerd
Daniel Recasens
Rosemarie Rotunno
Hyla Rubin
Judith Rubin
Arnold Shapiro
Suzanne Smith
Ben C. Watson
Douglas Whalen
Deborah Wilkenfeld
David Williams

---

*Part-time

[1] Scuola Normale Superiore, Pisa, Italy

[2] Visiting from University of Colorado, Boulder, Colorado

[3] Visiting from University of Tokyo, Japan

CONTENTS

I. Manuscripts and Extended Reports

Exploring the functional significance of physiological tremor: A biospectroscopic approach--David Goodman and J. A. Scott Kelso 1-27

Differences between experienced and inexperienced listeners to deaf speech--Nancy S. McGarr 29-51

A language-oriented view of reading and its disabilities--Isabelle Y. Liberman 53-75

Phonetic factors in letter detection: A reevaluation--Adam Drewnowski and Alice F. Healy 77-98

Categorical perception: Issues, methods, findings--Bruno H. Repp 99-183

Short-term recall by deaf signers of American Sign Language: Implications of encoding strategy for order recall--Vicki L. Hanson 185-203

A common basis for auditory sensory storage in perception and immediate memory--Robert G. Crowder 205-219

Phonological awareness and verbal short-term memory: Can they presage early reading problems?--Virginia A. Mann and Isabelle Y. Liberman 221-237

Initiation versus execution time during manual and oral counting by stutterers--Gloria J. Borden 239-254

Trading relations in the perception of speech by five-year-old children--Rick C. Robson, Barbara A. Morrongiello, Catherine T. Best, and Rachel K. Clifton 255-274

The role of the strap muscles in pitch lowering--Donna Erickson, Thomas Baer, and Katherine S. Harris 275-284

Phonetic validation of distinctive features: A test case in French--Leigh Lisker and Arthur S. Abramson 285-292

On consonants and syllable boundaries--Katherine S. Harris and Fredericka Bell-Berti 293-299

Vowel information in postvocalic frictions--D. H. Whalen 301-309

II. Publications 313

III. Appendix: DTIC and ERIC numbers (SR-21/22 - SR-69) 315

I. MANUSCRIPTS AND EXTENDED REPORTS

EXPLORING THE FUNCTIONAL SIGNIFICANCE OF PHYSIOLOGICAL TREMOR: A BIOSPECTROSCOPIC APPROACH

David Goodman+ and J. A. Scott Kelso++

Abstract. The functional significance of physiological tremor--the high frequency (8 to 12 Hz), low amplitude oscillation that occurs during the maintenance of steady limb postures--is not known. Often tremor--perhaps because of its pathological manifestations--is considered a source of unwanted noise in the system, something to be damped out or controlled. An examination of the phase relationship between tremor and rapid voluntary finger movement in normal subjects suggests a very different view. In four experiments in which tremor displacement and accompanying electromyographic activity were simultaneously monitored, we show a clear and systematic relationship between tremor and movement initiation. Empirically obtained frequency distributions of tremor peak-to-movement initiation time were most closely aligned to a probability density function (derived via numerical integration techniques) that assumed movements were initiated when the muscle-joint system possessed peak momentum. This relationship--evaluated by Chi-square goodness-of-fit tests--was evident regardless of whether the movements were self-paced (Experiments 1 and 3) or in response to an auditory reaction time signal (Experiments 2 and 4). The addition of a load to the finger in Experiments 3 and 4, though tending to reduce tremor frequency, did not prove disruptive, nor did a fractionated reaction time analysis reveal any significant inertial contribution to the maintenance of the phase relationship. These data are consistent with an emerging view that the motor control system is sensitive to its own dynamics, and suggest that under certain conditions normal physiological tremor is a potentially exploitable oscillation intrinsic to the motor system.

+Department of Kinesiology, Simon Fraser University, Burnaby, British Columbia.

++Also Departments of Biobehavioral Sciences and Psychology, University of Connecticut, Storrs, Connecticut.

Acknowledgment. A portion of this research was submitted in partial fulfillment of the Ph.D. degree to the University of Iowa by the first author. The work was performed at Haskins Laboratories and was supported by NINCDS Grant NS-13617, NIAMDD Grant AM-25814, and BRS Grant 05596. The authors would like to thank Charles Marshall, Tom Baer, Kiyoshi Honda, and David Zeichner for their help throughout the study.

[HASKINS LABORATORIES: Status Report on Speech Research SR-70 (1982)]

INTRODUCTION

Physiological tremor is a high frequency (in the 8 to 12 Hz range), low amplitude oscillation that occurs during the maintenance of steady limb postures. Although first described by Horsley and Schaffer in 1886, the origin and functional significance of "normal" tremor is still unclear today (Marsden, 1978; Stein & Lee, 1981). A number of candidates have been proposed as causes of tremor. One view is that tremor arises as a visco-mechanical property of each muscle load system (Randall, 1973; Rietz & Stiles, 1974). According to this hypothesis, normal tremor is thought to represent vibration caused by continuous broad frequency-band forcing of an underdamped, second order system at, or near, its natural frequency. Another possible source of tremor may be that produced by patterns of motoneuron discharge that occur when muscles contract (Sutton & Sykes, 1967). These can be further separated into three basic categories: First, the inherent firing properties of motoneurons per se; second, an instability in the stretch reflex arc associated with sychronization of motoneuron discharge at 8 to 12 Hz; and third, supraspinal rhythmic input to motoneurons (cf. Marsden, 1978, for review). Over the years some investigators have favored one source more than another. However, in spite of differences in emphasis, no single view as to the cause of physiological tremor has emerged, a view aptly summed up in Matthews and Muir's (1980) comment that: "After prolonged debate on the origins of physiological tremor, it is becoming increasingly accepted that tremor in the 8 to 12 Hz range may result from a variety of interacting mechanisms, one or other of which may predominate under any particular condition" (p. 429).

The present paper is not concerned directly with the causes of tremor, but rather addresses an equally intriguing--but less frequently considered--problem. What role, if any, does tremor play in the initiation and control of movement? It is fair to say that the general consensus on this issue is that tremor is a source of unwanted noise, something to be controlled rather than exploited. Such a view is evident, for example, in a preface to a recent volume dedicated to understanding the mechanisms of physiological tremor. Tremor is deemed as "...not useful...to have tremor oscillations cannot help by themselves, even indirectly, to make the motor performance faster or better" (Desmedt, 1978, p. vii). Consonant with this perspective, currently popular closed-loop, servomechanism models of motor behavior--with their emphasis on set points and error correction processes--consider oscillatory behavior a nuisance, an unwanted source of variability (e.g., Adams, 1971).

Given the existence of cyclicities operating at many different levels in biological systems, it may be premature (if not myopic) to reject a functionally significant role for oscillation in general, and physiological tremor in particular. For example, many years ago Brown (1914) argued that rhythmic signals arising from oscillatory networks in the spinal cord were one of the foundations of integrative activity in the mammalian nervous system. Although this idea received rather spasmodic attention over the years, it is now becoming recognized as a fundamental insight (Delcomyn, 1980; von Holst, 1973). The potential importance of oscillatory processes in motor control is suggested not only by recent empirical investigations in the physiology of movement (Grillner, 1975; Shik & Orlovsky, 1976; Stein, 1976), but also by recent theoretical work in the emerging field of physical biology. Iberall (1972), for example, has characterized biological systems as ensembles of coupled and mutually entrained oscillators; stable organization, according to

Iberall's physical theory of homeokinesis, is a consequence of the interaction of oscillatory processes at all levels of the system. Cyclicity, in the homeokinetic view, is not some epiphenomenal property of biological systems; instead, all persistent, self-sustaining mechanisms (including living things) exhibit dynamic stability by virtue of nonlinear, limit cycle processes (cf. Iberall, 1977; Soodak & Iberall, 1978; Yates, 1979; Yates, Marsh, & Iberall, 1972). Rather than being viewed as an incidental aspect of biological systems, oscillatory behavior may be a central feature of their organization (Goodwin, 1970).

The approach that we adopt to the problem of tremor in this paper is that of "biospectroscopy"--the identification of cyclicities and determination of their functional significance--advocated by homeokinetic theory (for particular application to motor control and coordination see Kugler, Kelso, & Turvey, 1980, 1982, and for empirically related work see Kelso, Holt, Kugler, & Turvey, 1980; Kelso, Holt, Rubin, & Kugler, 1981). If it is accepted that oscillation is a fundamental dynamic property of living systems, then it seems possible that tremor is present for a reason and that under certain conditions, humans may actually use tremor to enhance motor performance. From mechanics we know that a system in continuous oscillation provided with an appropriately phased forcing function requires less energy to move than a system in static equilibrium. Is it possible then, that a systematic phase relationship exists between the initiation of movement and physiological tremor? An early study by Travis (1929)--not to our knowledge referred to in recent reviews of the tremor literature--hints strongly at such a possibility. Travis (1929) observed that a large proportion of upward movements were initiated during the ascending phase of tremor. Similarly, downward movements appeared to be produced during the descending phase of tremor. However, in order to examine the relationship (if indeed one exists) over a wider range of conditions, and to determine the locus on the tremor cycle around which voluntary movements may be initiated, a quantitative approach seems warranted.

In the present set of experiments, subjects were required to maintain a steady, stable position of the index finger while tremor and electromyographic activity from the primary extensor were simultaneously monitored. In Experiments 1 and 2, subjects initiated upward ballistic movements of the index finger in a self-paced manner, or under time stress conditions in response to an auditory stimulus. The time stress experiment (basically a simple reaction time situation) was included to determine if inducement to respond as quickly as possible would override the hypothesized phasing between movement onset and tremor. The self-paced and time-stressed paradigms were used in two further experiments in which a load was also added to the finger in order to increase the inertia of the muscle joint system. By fractionating movement initiation time into its so-called premotor (latency of signal onset to EMG onset) and motor (latency of EMG onset to movement onset) components (cf. Botwinick & Thompson, 1966; Weiss, 1965) we sought to evaluate a possible inertial contribution to the phase relationship. That is, a relationship between peripheral motor time and movement initiation time would suggest that mechanical lag in the muscle-joint system contributes significantly to the phasing.

THE MODELS

Four models were generated according to different assumptions about the time of voluntary movement initiation with respect to the physiological tremor cycle (measured as a peak-to-peak time interval). All the models used the conjoint distribution of tremor peak-to-peak times and peak-to-movement initiation times (obtained from displacement-time records) to derive probability density functions. Numerical integration was used to compute the four theoretical distributions that were then compared to the actual distribution of peak-to-movement initiation times obtained from the data.[1] The details of the derivation of each model are provided in Appendix 1; Figure 1 shows the actual theoretical distributions.

Model 1 postulates no systematic relationship between the initiation of movement and physiological tremor. The probability of movement initiation is therefore uniformly distributed throughout the peak-to-peak interval, and may be described by the following probability density function:

$$f(y) = \int_{y}^{\infty} \left[\frac{1}{x} I_{[o,x]}(y)\right] \left[\frac{1}{\sqrt{2\pi}\, S_x} e^{-1/2 \frac{(x-\bar{x})^2}{S_x^2}}\right] dx \tag{1}$$

where x is a random normal variable of tremor peak-to-peak time, x is the sample mean, S_x^2 is the sample variance, y is a random variable defined as peak-to-peak movement initiation time, and $I_{[o,x]}(y)$ defines the interval for the uniform distribution of y.

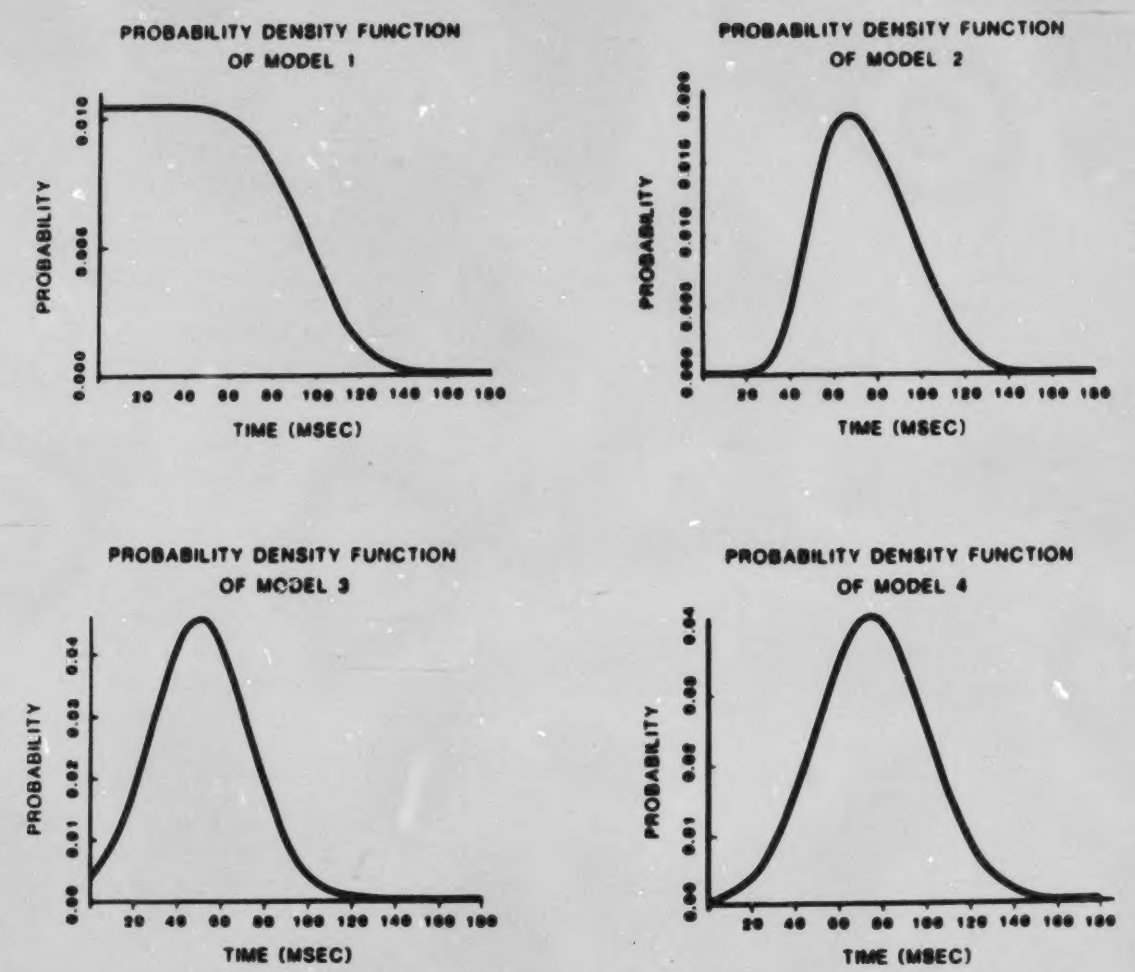
Model 2 assumes that the initiation of upward movement is equally dispersed throughout the ascending phase of the tremor. Thus the probability of movement initiation may be uniformly distributed throughout the ascending phase (from trough to peak), describable by the following probability density function:

$$f(y) = \int_{y}^{2y} \left[\frac{2}{x} I_{[\frac{1}{2}x,x]}(y)\right] \left[\frac{1}{\sqrt{2\pi}\, S_x} e^{-1/2 \frac{(x-\bar{x})^2}{S_x^2}}\right] dx \tag{2}$$

Model 3 assumes that the forcing function is applied when the muscle-joint system possesses maximum potential energy. Since the potential energy of an oscillatory system is proportional to its displacement, the point of maximum potential energy for an upward movement is at the trough of the tremor cycle. Hence the probability density function has the following form:

$$f(y) = \int_{x=y}^{\infty} \left[\frac{1}{\sqrt{2\pi}\, S_x} e^{-1/2 \frac{(x-\bar{x})^2}{S_x^2}}\right] \left[\frac{1}{\sqrt{2\pi}\, S_y} e^{-1/2 \frac{(y - \frac{\bar{x}^2}{2})}{S_y^2}}\right] dx \tag{3}$$

Model 4 follows from a minimum energy hypothesis in which the forcing function is applied when the system possesses peak momentum. Since momentum is proportional to mass and velocity, and since mass is held constant in this case, the point of maximum momentum is at the inflection point of the upward


PROBABILITY DENSITY FUNCTION OF MODEL 1

PROBABILITY DENSITY FUNCTION OF MODEL 2

PROBABILITY DENSITY FUNCTION OF MODEL 3

PROBABILITY DENSITY FUNCTION OF MODEL 4

Figure 1. Probability density functions derived from theoretical distributions based on different assumptions regarding the phase relationship between voluntary movement initiation and physiological tremor (see text and Appendix 1 for details).

phase of the tremor cycle. Therefore the probability density function takes the following form:

$$f(y) = \int_{x=y}^{\infty} \left[\frac{1}{\sqrt{2\pi}\; S_x} \; e^{-1/2 \; \frac{(x-\bar{x})^2}{S_x^2}} \right] \left[\frac{1}{\sqrt{2\pi}\; S_y} \; e^{-1/2 \; \frac{(y - \frac{3}{4}\bar{x})^2}{S_y^2}} \right] dx \qquad (4)$$

METHODS

Subjects. Each experiment was limited to three subjects. The same three subjects served in Experiments 1 and 2; a different three subjects served in Experiments 3 and 4. The subjects were adult male volunteers who were not compensated for their participation. All subjects signed informed consent forms that described the experiments and any accompanying risks and benefits. Subjects were free to withdraw their participation at any point if they so chose.

Apparatus. A linear variable differential transducer (LVDT, Model PCA 116-100, Schaevitz) 5.0 cm long by 2.1 cm in diameter, was mounted in an adjustable wooden arm such that the transducer was suspended over and above the extended finger of the subject. A 2.0 cm diameter wooden dowel served as a hand grasp and was mounted horizontally 7.6 cm above a standard height table, 12.7 cm from the table's leading edge and parallel to it.

The LVDT was coupled to an amplifier, and the resultant signal displayed on an oscilloscope and stored on FM tape. The transducer was able to detect movements as small as .025 mm, while the actual weight resting on the fingertip was approximately 10 grams. An oscilloscope was positioned behind the table at eye level, directly in the field of vision of the subject. Two horizontal bars, centered 4 cm apart on the oscilloscope display screen served to define the acceptable field of movement. Bipotential hooked-wire electrodes were used to obtain electromyographic (EMG) signals from the extensor digitorum communis. In Experiments 2 and 4 a Minisonalert (Mallory) was employed to generate an auditory stimulus. The Minisonalert was situated approximately 1 meter in front of the subject and generated a high-pitched tone (approximately 2900 Hz) for a duration of 8 msec upon switch closure by the experimenter. In Experiments 3 and 4 a 200 gm metal disk (a 100 gm disk was used by one of the subjects who had difficulty initiating movements with the heavier disk) of 4.2 cm diameter was taped under the distal phalangeal joint of the index finger. The load itself did not interfere with the range of motion.

Procedures. The same general procedure was employed in all four experiments. Specific procedures are detailed only insofar as they deviate from those described below. In preparation for the insertion of EMG electrodes the subject sat in a chair facing the experimental table. Bipolar, hooked-wire electrodes consisting of a pair of platinum-tungsten alloy wires (50 microns in diameter with isonel coating) were inserted into the extensor digitorum communis by means of a 26 gauge hypodermic needle. Before insertion, subcutaneous anesthesia (1% Xylocaine) was applied to the area of insertion by means of a Panjet injector. For verification of electrode position, the subject performed flexion and extension movements of the right index finger about the metacarpophalangeal joint; during these maneuvers the EMG signals were monitored on an oscilloscope and over a loudspeaker. After amplification and high-pass filtering at 80 Hz to remove movement artifacts and hum, the signals were recorded on a multichannel instrumentation tape. The signal from the displacement transducer was simultaneously recorded. The subject placed his right arm on the table, grasped the wooden dowel, then extended the right index finger and maintained it in a horizontal position. The wooden arm supporting the linear transducer was then adjusted so that the transducer was positioned directly above the center of the fingernail of the extended finger. The mid-range position of the finger was associated with a straight line tracing on the oscilloscope, centered between the two horizontal bars.

Each experiment proceeded through an initial practice session followed by the experimental session. The practice session consisted of as much time as needed for the subject to establish a sufficiently stable tremor to allow the recording session to proceed. The subject was instructed to watch the oscilloscope tracing and to maintain the position of the tracing between the two horizontal bars on the screen as well as possible. Approximately 10 min of practice were usually necessary. For all experiments the subject was required to maintain a stable position for approximately 2 sec and then produce a rapid upward movement of the index finger. While the movement itself was to be made rapidly, the time of onset was either self-paced or in response to an auditory stimulus, dependent on the particular experimental manipulation. Experiments 1 and 3 were self-paced; that is, the subject initiated the movements at his own pace. In Experiments 2 and 4 the movements were made as rapidly as possible following the onset of an auditory stimulus. The time of onset of the stimulus was controlled by the experimenter. After making the movement, the subject returned the finger to the mid-range position, held it stable for a short time and then repeated the sequence a total of 200 times. A 20 sec rest was given after each set of ten trials and a two minute rest after the fiftieth, one hundredth, and one hundred and fiftieth trials. The subject was permitted as much time as necessary to stabilize the finger between each trial and additional rest periods were taken as needed.

Data analysis. An analogue to digital conversion was made by reading simultaneously from the two channels (displacement and EMG) on the FM tape and saving the digital conversion in direct access files. Each signal was sampled at 5 kHz and low-pass filtered at the Nyquist limit. The displacement signal was downsampled and smoothed by means of a monotonic low-pass filter to remove frequencies over 30 Hz. The electromyographic signal, which was time locked to displacement, was rectified and integrated into 5 msec bins. A wave editing and display routine (WENDY; Szubowicz, Note 1) was used to display and label each record as shown in Figure 2. In Figure 2, PK corresponds to the last clearly defined peak of tremor before the upward movement; MO defines the time of movement onset, as indicated by the displacement curve going off scale. Note that this is necessarily an overestimate (approximately 12 msec on the average); and EM is the time of the first EMG activity associated with upward movement as indicated by the onset of the initial rise of activity on the rectified and integrated EMG record. In addition, in Experiments 2 and 4, the onset of the auditory stimulus was labeled as RT. The latency from the signal to EMG onset allowed for the determination of so-called premotor time, and the latency of EMG onset to movement onset was indicative of the motor component of reaction time (cf. Botwinick & Thompson, 1966; Weiss, 1965).

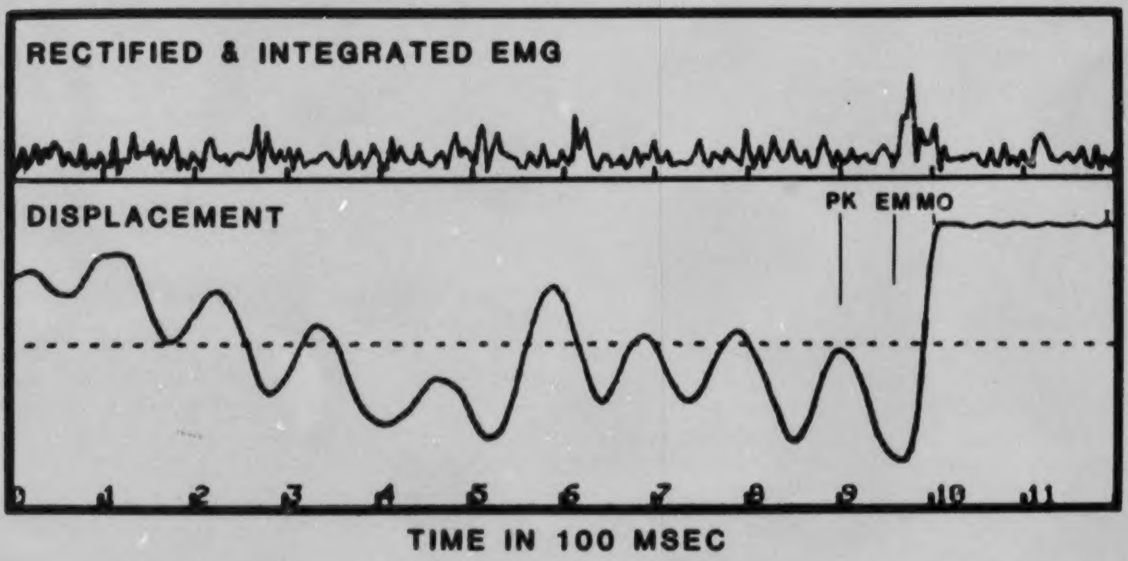


Figure 2. Sample record of tremor displacement-time profile and associated electromyographic activity. Marker labels defined as in text.

Although each subject made 200 movements in each of the four experiments, the number of trials included in the final analysis was lower due to the rigorous conditions for retention of a trial. The most frequent reasons for rejection of a trial were either that there were not two clearly defined peaks of tremor just prior to movement initiation, or that the displacement record was out of range of the measuring instrument. A less frequent reason for rejection was that the EMG record was of poor quality. In addition, in Experiments 2 and 4, trials in which the reaction times were less than 70 msec or greater than 600 msec were rejected. This is a standard procedure used in reaction time studies to reduce the respective effects of anticipation and inattention (cf. Goodman & Kelso, 1980).

In order to determine the best fitting theoretical distribution, a linear transformation was made so that the data could be collapsed over all subjects. Each individual subject's data were transformed such that the last peak-to-peak interval before movement onset (peak n-1) had a mean of 100 msec and standard deviation of 20 msec. This mean value is consistent with the literature, representing a tremor oscillation of 10 Hz, and the standard deviation was empirically determined from pilot data. Each of the four theoretical models was based on tremor peak-to-peak times with the above distribution. The transformed data were then analyzed in a similar manner to the individual subject data to produce a frequency distribution, mean, and standard deviation. These resulting distributions for each of the four experiments were compared to the four theoretical distributions by means of Chi square goodness-of-fit test.

Results and Discussion

Three aspects of the results are presented in turn. First, a reliability analysis on the measurements of interest is given followed by a summary analysis for all experiments. The last section deals with tests of the four theoretical models.

Reliability of measures. We first conducted a reliability check on the main measures of interest, namely, the movement onset and the EMG onset. Every fourth trial of a randomly chosen subject's (S3) performance was measured a second time by a person not familiar with the purposes of the investigation. This second "measurer" was instructed to label each of the movement records given only the definition of each event as described in the previous section (i.e., PK, MO, and EM). These data were tabulated in the same manner as the originally measured data and were then correlated. For movement onset the mean difference between measures was 2.7 msec. The high reliability was not totally unexpected, given the rigorous conditions for retention of a trial. For EMG onset the mean difference was 1.3 msec. The reliability coefficient exceeded 0.90 for both dependent measures.

Experiment 1. The first experiment involved self-paced movements without load, the results of which are summarized in Table 1. All subjects had a tremor rate ranging between 9.1 and 10.2 Hz, which is consistent with previous estimates (e.g., Rack, 1978). The variability of the tremor cycle-to-cycle time was considerable, with an average standard deviation across subjects of 17.4 msec. The time of movement onset was approximately 90% of the way through the tremor cycle. It should be emphasized again, however, that the method of measuring movement onset time was necessarily a slight overestimate.

Table 1

Means (and Standard Deviation) in Msec for Each of the Subjects in Experiment 1 (Self-paced, Unloaded)

	Variable		
Subject	Peak-to-Peak time[a]	Peak-to-EMG Onset[b]	Peak-to-Movement onset
1	98.0 (20.1)	50.5 (24.7)	95.0 (26.8)
2	109.5 (14.0)	34.8 (29.8)	92.9 (27.7)
3	101.9 (18.1)	34.3 (26.3)	91.7 (26.7)

[a] Interval between last two measured peaks before movement onset

[b] As measured from rectified and integrated signal

The correlation between the onset of the rectified and integrated EMG signal and movement initiation, as defined here, was quite high (r=.84) and the average lag between these variables was 53 msec which is again consistent with other data (e.g., Desmedt & Godaux, 1978).

Experiment 2. In Experiment 2 subjects responded as quickly as possible to an auditory signal by making an upward movement of the index finger as quickly as possible. As shown in Table 2, the tremor rate was similar to Experiment 1 (8.8 Hz to 9.7 Hz), with an average standard deviation in periodicity of 18.4 msec. Time of peak-to-movement onset was, as in Experiment 1, approximately 90% of the tremor cycle time.

The results of the fractionated reaction time analysis are also given in Table 2. The reaction time (mean of 258 msec) was highly correlated to premotor time (mean of 208 msec; r = .97) while uncorrelated with motor time (mean of 49 msec; r < .01). The partial correlation of motor time to total reaction time (with the variance of reaction time due to premotor time parceled out) was negligible (r < .01). The independence of premotor time and motor time (r = -.174) was consistent with that reported by others (Botwinick & Thompson, 1966), which also showed little or no correlation between these variables.

Experiment 3. In Experiment 3 subjects produced self-paced movements with a load added to the finger. The results of the overall experiment are summarized in Table 3. Although cross-experimental comparisons are tenuous, it appears that the addition of load reduced the tremor rate in two of the subjects. The remaining subject had only a 100 g load attached to the appendage, and his tremor rate was well within the bounds of normal physiological tremor. These data suggest that heavier loads are associated with reduced tremor rate, a notion not inconsistent with other findings showing that increasing the moment of inertia of the vibrating part reduces frequency of oscillation (Stiles & Randall, 1967). On the other hand, there are data showing no change in finger tremor rate with added mass of up to 100 gm (Halliday & Redfern, 1958).

Time of tremor peak-to-movement onset was similar to that observed in Experiments 1 and 2 for two of the subjects (for S1 and S2, 86% and 91% of the cycle time, respectively). For the third subject, however, movements tended to be initiated earlier in the cycle (51% of the cycle time). The correlation between movement initiation and onset of EMG was again quite high (r = .88) with a lag time of 68 msec. This slight increase in lag time, compared to Experiments 1 and 2, is not unexpected because adding a load is likely to prolong the mechanical contractile latency of muscle (cf. Desmedt & Godaux, 1978, for review).

Experiment 4. The results of Experiment 4, in which subjects produced movements under loaded conditions as rapidly as possible following the onset of an auditory stimulus, are given in Table 4. Tremor rates remained somewhat slower than normal (as compared to Experiments 1 and 2) for two of the subjects (although S3 had an increased rate of tremor, 7.8 Hz, compared to Experiment 3). The relative time of movement onset with respect to the tremor cycle was again similar to that of Experiment 3 (81.5% to 86%). The results of the fractionated reaction time analysis are also given in Table 4. The reaction time (mean of 264 msec) was correlated to pre-motor time (mean of 193

Table 2

Means (and Standard Deviation) in Msec for Each of the Subjects in Experiment 2 (Reaction Time, Unloaded)

Subject	Peak-to-Peak time[a]	Peak-to-EMG onset[b]	Peak-to-Movement Onset	Reaction Time	Premotor Time	Motor Time
			Variable			
1	113.7 (21.2)	46.4 (30.8)	97.9 (27.4)	268.9 (55.7)	217.5 (50.9)	51.4 (17.8)
2	108.7 (15.5)	42.5 (25.2)	85.1 (49.9)	241.1 (53.0)	198.4 (50.9)	42.6 (11.2)
3	102.8 (18.6)	31.7 (19.7)	86.2 (18.9)	264.5 (35.2)	210.0 (35.4)	54.5 (8.7)

---

[a] Interval between last two measured peaks before movement onset.

[b] As measured from rectified and integrated EMG

Table 3

Means (and Standard Deviation) in Msec for Each of the Subjects in Experiment 3 (Self-paced, Loaded)

	Variable		
Subject	Peak-to-Peak time[a]	Peak-to-EMG Onset[b]	Peak-to-Movement onset
1	170.1	59.1	146.8
	(59.6)	(56.6)	(62.0)
2	112.5	32.5	102.9
	(28.9)	(39.4)	(36.0)
3	208.1	62.0	107.7
	(63.4)	(37.5)	(34.0)

---

[a] Interval between last two measured peaks before movement onset

[b] As measured from rectified and integrated signal

Table 4

Means (and Standard Deviation) in Msec for Each of the Subjects in Experiment 4 (Reaction Time, Loaded)

	Variable					
Subject	Peak-to-Peak time[a]	Peak-to-EMG onset[b]	Peak-to-Movement Onset	Reaction Time	Premotor Time	Motor Time
1	155.6 (29.6)	15.3 (30.8)	95.8 (19.6)	298.6 (46.0)	218.1 (37.0)	80.6 (16.0)
2	107.5 (18.3)	4.5 (30.8)	95.3 (27.8)	287.6 (47.8)	196.8 (48.6)	90.8 (16.2)
3	127.6 (30.1)	69.0 (30.1)	110.3 (27.5)	206.5 (44.6)	165.3 (44.4)	41.3 (11.6)

---

[a] Interval between last two measured peaks before movement onset

[b] As measured from rectified and integrated EMG

Table 5

Expected Cumulative Frequency (in percent)
for the Four Theoretical Distributions

	Theoretical Distribution Derived From			
Frequency Bounds (upper limit)	Model 1	Model 2	Model 3	Model 4
25	26	0	12	2
35	37	1	24	6
45	47	6	41	12
55	58	18	58	21
65	67	35	74	35
75	77	53	85	50
85	85	69	93	66
95	91	82	95	73
105	97	91	100	88
115	100	100	100	95
125	100	100	100	98
>125	100	100	100	100

Table 6

Actual Cumulative Frequency (in percent) for the Four Experiments

Frequency Bounds (upper limit)	Actual Distributions			
	Exp 1	Exp 2	Exp 3	Exp 4
25	2	5	0	0
35	6	8	0	0
45	13	13	0	9
55	19	20	6	18
65	26	40	22	32
75	42	46	43	39
85	49	57	60	61
95	61	71	72	74
105	71	83	79	86
115	84	87	90	95
125	89	97	97	96
>125	100	100	100	100
N[a]	149	87	206	57

[a] Actual number of observations

msec, $r = .94$), and uncorrelated to motor time (mean of 71 msec, $r = -.02$). As in Experiment 2, the partial correlation of motor time to total reaction time was negligible ($r < .01$). This result concurs with other investigators (Kamen, 1980) who have found reaction time to be related to premotor time but not motor time in both unresisted and resisted cases.

Test of models. The basic question of interest in all the experiments was the existence and nature of the phase relationship between the initiation of movement and physiological tremor. Analysis of each separate experiment produced a frequency distribution that allowed for a comparison with each of the four theoretical models. Thus each experiment, while analyzed separately, was treated similarly with respect to the above question. The number of movement onsets within each 10 msec interval and the consequent frequency distributions generated are shown for each of the four experiments in Figure 3. Table 5 gives the expected cumulative proportion for those same intervals, derived from each of the theoretical distributions, and Table 6 gives the actual cumulative proportion derived from each of the experiments. A summary table of Chi-square goodness-of-fit tests is presented in Table 7, and indicates a similar pattern for the four experiments. That is, the Chi square goodness of fit was smallest when the empirical distributions obtained from each of the four experiments were compared to the theoretical distribution of Model 4. This result alone suggests that the initiation of voluntary movement is not arbitrary with respect to tremor, but rather occurs systematically in phase with it.

Table 7

Chi Square Goodness of Fit Tests (and Degrees of Freedom)
Between Empirical Distributions from the Four Experiments
and the Theoretical Models

Experiment	Model 1	Model 2	Model 3	Model 4
1	262.6 (17)	196.0 (17)	566.2 (15)	65.6 (19)
2	111.4 (17)	113.0 (17)	155.8 (15)	38.4 (19)
3	297.1 (12)	104.7 (14)	553.4 (10)	67.5 (15)
4	79.4 (14)	20.3 (16)	118.2 (17)	18.7 (17)

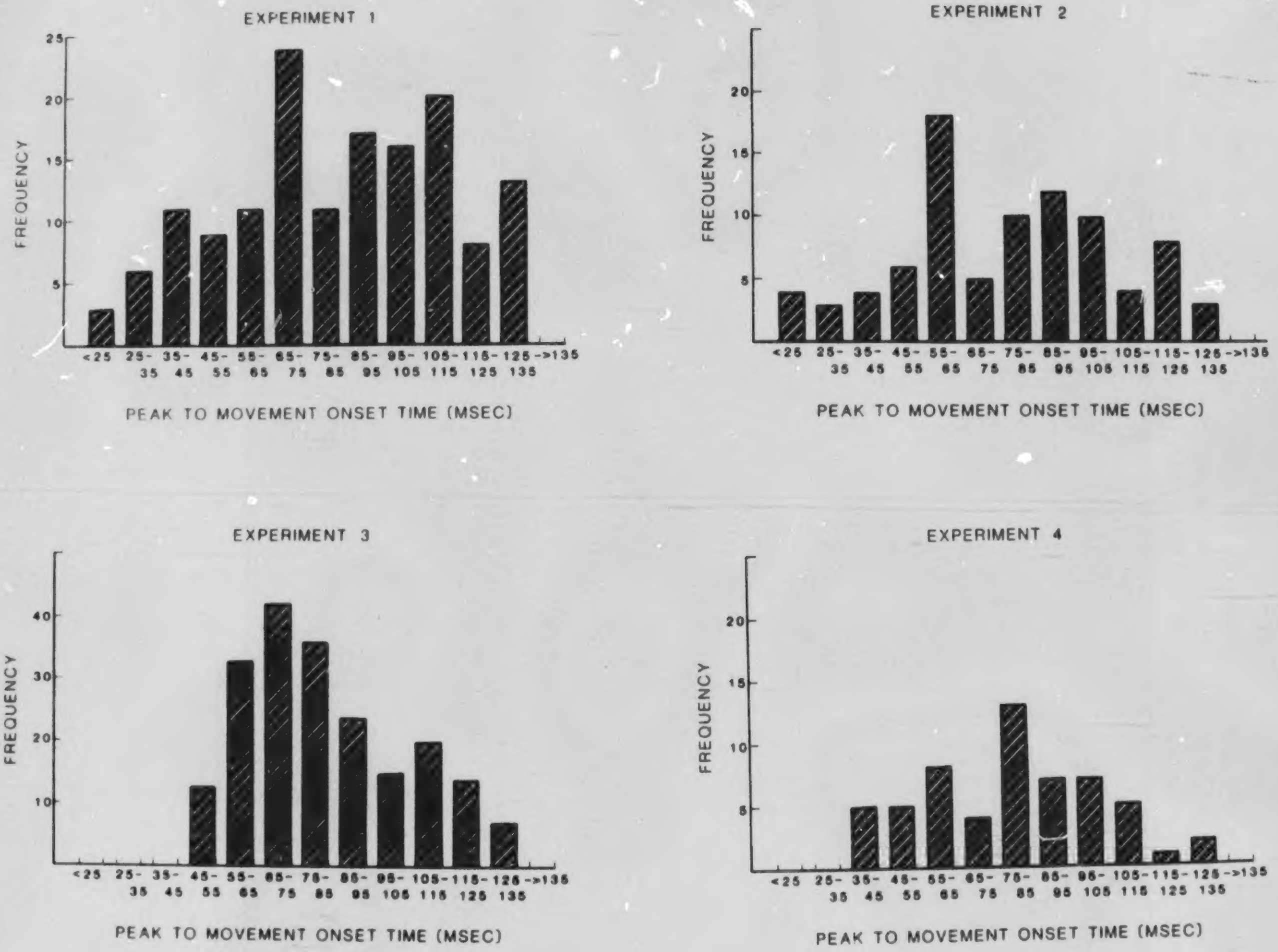


Figure 3. Frequency distributions of transformed peak-to-movement initiation times for all four experiments. Experiments 1 and 2 correspond to self-paced and reaction time conditions for unloaded movements. Experiments 3 and 4 involve the same conditions but with a load attached to the finger.

Additional support for the foregoing claim is provided by the large Chi square obtained by comparing the empirical distributions to the theoretical distribution of Model 1. Had there been no relationship between movement initiation and physiological tremor, a model based on movements occurring with equal probability throughout the tremor cycle would have been supported. Such was not the case: in each experiment the resultant Chi square for Model 1 was over three times as large as the Chi square obtained for Model 4. Model 3 can also be rejected on these grounds for each of the experiments.

The distinction between Model 2, which postulates a simple phase relationship between movement initiation and physiological tremor, and Model 4, which postulates a more exact relationship between the two variables, is not quite as clear, particularly when the appendage was loaded (Experiments 3 and 4, see Table 7). However, in all cases Model 4 had a lower Chi square than Model 2 (sometimes by a factor of 3) and therefore appears the most likely candidate.

Neither is there evidence to support the notion that the phase relationship between physiological tremor and movement initiation breaks down when a premium is placed on responding quickly. In support of this claim are the small Chi squares obtained for Model 4 in both of the experiments requiring a speeded response (Experiments 2 and 4). Although in all experiments there was a small proportion of trials in which subjects initiated a response that was not in phase with the tremor cycle (as reflected in the tails of the distributions in Figure 3), this proportion remained relatively constant across experimental conditions.

In summary, the data from all four experiments show a strong tendency for upward ballistic movements to be initiated in the upward phase of the tremor cycle. Moreover, the point of initiation appears to be distributed around the point in the tremor cycle at which the muscle-joint system possesses peak momentum.

GENERAL DISCUSSION

Cyclicities in biological systems have been long established in the literature and range in periodicity from years, as in predator-prey cycles, to months, as in the menstrual cycle, to days, as in circadian phenomena, to fractions of seconds, as in certain neural events. One of these cyclicities, and the subject of investigation in the present paper, is physiological tremor. Tremor has intrigued physiologists and clinical neurologists for a long time, with most of the research effort targeted to questions regarding its origin. What generates tremor? Even Travis (1929) whose work first hinted that "...willed movement is not independent of the tetanic (tremor) contractions...but blends into the rhythm already established" seemed preoccupied with the question of where tremor came from. Without any evidence to speak of, Travis postulated that physiological tremor and voluntary movement had common origins in the cerebral cortex. As we noted in the introduction to this article, answers to the question of origins, however, still remain elusive (cf. Marsden, 1978; Stein & Lee, 1981).

Sidestepping the origins issue, the present experiments were directed to an issue of equal puzzlement to physiologists, namely, the functional significance of normal, physiological tremor. Travis's (1929) early work, along with recent theoretical considerations that oscillatory processes play a central, organizing role in complex systems with many degrees of freedom (Iberall, 1972; Soodak & Iberall, 1978; Yates et al., 1972; see also Kelso, 1981; Kugler et al., 1980, 1982, for applications to movement control issues), suggest that oscillations are present for a reason. The intuition is that living systems may be designed to take advantage of intrinsic oscillatory processes.

As far as the control of movement is concerned, it seemed possible that tremor may be used as a type of background facilitation for voluntary movement. The four experiments reported here offer strong support for the notion that tremor is exploitable. In all cases, we observed a systematic phase relationship between movement initiation and tremor. Moreover, movement initiation appeared to be distributed around the point at which the muscle-joint system possessed peak momentum (Model 4).

The present results are consistent with a general theme that is only recently receiving its due notice; namely, that the motor control system is sensitive to its own physical dynamics and is capable of taking advantage of them (Cooke, 1980; Greene, 1972; Kelso, 1981; Kelso & Holt, 1980; Kelso et al., 1980; Kugler et al., 1980, 1982). With respect to the findings here, it is noteworthy that kinetic energy is greatest around the point of maximum momentum in an oscillation. Presumably, if the motor system was "smart" (paralleling Runeson's [1977] smart perceptual device), it would take advantage of this fact for reasons of energy optimization. In short, it would be cost-efficient for voluntary movement initiations to be distributed around the point of peak momentum (maximum angular velocity). Note that in order to initiate movement around this point, the mechanical lag between onset of electromyographic activity and movement must be taken into account. That this appears to be so in the present experiments suggests that the nervous system is sensitive to the physical facts of oscillation. There is, as it were, a mutual coupling between the information, signalling aspects, and the power plant provided by muscles.

That a highly evolved system may take advantage of intrinsic oscillations for the purpose of reducing the energy demands associated with movement, is supported by studies that measure the energy requirements of sustaining sinusodial movements of a limb. Rack and his colleagues coupled the elbow joint to a machine capable of driving the joint sinusoidally and found that below 6 Hz and above 13 Hz the machine had to do work to sustain the movement; however, between 6 and 13 Hz (peaking around 10 Hz) the limb actually did work on the machine (cf. Rack & Westbury, 1974). Thus the amount of energy required to drive the limb at its natural resonant frequency (coinciding with tremor) was much less than at other frequencies (see Rack, 1978, Figures 4 and 5). Although Rack's findings are consistent with the present data and help to rationalize them, they do not address the issue germane to the present studies, viz., the phasing of volitional activity and tremor.

The results of the experiments reported here are particularly relevant to the work of a group of Russian investigators (cf. Aizerman & Andreeva, 1968; Chernov, 1968). In a series of studies this group provided qualitative evidence that when the arm is held in a particular position, opposing agonist-antagonist muscles alternately pull the arm one way and then the other, producing a "tremor" of about ten cycles per second. The EMG envelopes of both muscles in the Soviet studies were observed to display "peaks" that appeared to arise each time the absolute value of joint angle velocity reached a certain threshold value. These peaks alternate in that if at one moment the peak is large for the flexor and small for the extensor, the next time the threshold value is reached, a large peak is observed for the extensor and a small one for the flexor. In this way movements in one direction or another are associated with increases in the amplitude of the EMG tremor peak of the involved muscle. In Aizerman's model, the brain is envisioned as sending the same signals to each muscle contributing to the limb's movement at the tremor frequency, while prior adjustments in the interneuronal pools allow each muscle to respond by the appropriate amount.

Our findings, suggesting that movement initiation is distributed around the point of peak angular momentum, fit rather well with Aizerman's "threshold" concept in which "splashes" of neuromuscular activity occur in relevant muscles when the joint reaches a critical angular velocity. Moreover, the idea that there may be critical values of certain system-sensitive parameters (or in the background state of interneuronal pools) that establish optimum conditions for control, receives support in larger scale activities such as human handwriting. In an elegant model of cursive handwriting that uses coupled oscillations in horizontal and vertical directions to produce letter forms, Hollerbach (1978) has shown that letter height modulation is best accomplished by altering acceleration amplitude at the vertical zero crossing. This point occurs at the top and bottom of letter corners and, in terms of the present study, would be associated (roughly) with the onset of EMG activity observed in the present experiments.

The present data also offer an empirical basis for the more recent speculations of Hallett, Shahani, and Young (1977) on Parkinson patients, that "...some of the delay in initiating movement in patients with tremor-at-rest might come from 'waiting to get into the correct time of the cycle'..." (p. 1133). Our results concur and suggest that the "correct time" may be distributed around a point at which it is physically advantageous to initiate movements.

That there appears to be value in having a low level oscillation in the limb segment before movement initiation, and that the cycling activity is exploited in energetically useful ways, is a claim in sharp contrast to conventional views of physiological tremor. Up to now, and possibly because of a preoccupation with pathological tremors, physiologists have tended to consider low-level oscillations as unwanted sources of noise. Similarly, physiological tremor is posited to occur "as a result of instability in the servomechanism associated with the spinal stretch reflex" (cf. Stein & Lee, 1981). The theoretical emphasis on "instability" and on ways to reduce tremor oscillations may have desensitized physiologists to the possible uses of tremor.

Tremor, as currently understood, is a stochastically fluctuating, quasi-periodic activity common to all humans, and is not, judging by present data, a source of "noise" in the conventional, undesired sense. Tremor "noise" appears to have a function, which may be to keep the system in motion in order to minimize its inertia and increase the velocity of its reactivity (Sollberger, 1965). As pointed out some years ago by Greene (1972), using small, rapid oscillatory movements might allow for graded control ("proportional" control) to be exerted by highly nonlinear and discontinuous systems. For example, a rapid fluctuating signal (or dither) added to a slowly varying control signal is often useful to overcome a threshold or "unstick" friction. The present data, as well as those discussed above (Aizerman & Andreeva, 1968; Hallet et al., 1977) can be interpreted as support of this view.

From a more general perspective, it is worth noting that physical biologists have recently discounted static, snapshot views of biological systems, in which the methodology dictates that periodic events are ignored (see Katchalsky, Rowland, & Blumenthal, 1974; Iberall, 1972). For persistence of function, living systems must conduct energy transactions in a cyclical manner if thermodynamic strictures are to be met. Such cycling is a general and inevitable consequence of the physics of open systems that undergo energy flux (Morowitz, 1979). Moreover, fluctuations in a system, according to contemporary physical theory, are a necessary precondition for the evolution and maintenance of function (Iberall, 1977, 1978; Prigogine, 1980). Extrapolating from such considerations, physiological oscillations in normal systems are not likely to be functionally insignificant.

In conclusion, the present data underscore the importance of giving oscillatory processes a more prominent role in our considerations of how movements are initiated and controlled. The findings reported here are consistent with evolving oscillator-theoretic views of neural control (cf. Delcomyn, 1980), and point to the gains that might be achieved when neuroscience and psychology embrace more fully design principles based on oscillatory processes.

REFERENCE NOTE

1. Szubowicz, L. *A tutorial guide to WENDY: The Haskins wave editing and display system*. Unpublished manuscript, Haskins Laboratories, 1977.

REFERENCES

Adams, J. A. A closed-loop theory of motor learning. *Journal of Motor Behavior*, 1971, *3*, 111-150.

Aizerman, M. A., & Andreeva, E. A. Simple search mechanism for control of skeletal muscles. *Automation and Remote Control*, 1968, *80*, 103-108.

Botwinick, J., & Thompson, L. W. Premotor and motor components of reaction time. *Journal of Experimental Psychology*, 1966, *71*, 9-15.

Brown, T. G. On the nature of the fundamental activity of the nervous centres; together with an analysis of the conditioning of rhythmic activity in progression, and a theory of evolution of function in the nervous system. *Journal of Physiology*, 1914, *48*, 18-46.

Brumlik, J. On the nature of normal tremor. *Neurology*, 1962, *12*, 159-179.

Chernov, V. I. Control over single muscles or a pair of muscle antagonists under conditions of precision search. *Automation and Remote Control*, 1968, *29*, 1090-1101.

Cooke, J. D. The organization of simple, skilled movements. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North-Holland Publishing Company, 1980.

Delcomyn, F. Neural basis of rhythmic behavior in animals. *Science*, 1980, *210*, 492-498.

Desmedt, J. E. *Physiological tremor, pathological tremor and clonus: Progress in clinical neurophysiology* (Vol. 5). New York: Karger, 1978.

Desmedt, J. E., & Godaux, E. Ballistic skilled movements: Load compensation and patterning of the motor commands. In J. E. Desmedt (Ed.), *Cerebral motor control in man: Long loop mechanisms*. New York: Karger, 1978.

Goodman, D. *The relationship between physiological tremor and human movement initiation*. Unpublished doctoral dissertation, University of Iowa, Iowa City, 1981.

Goodman, D., & Kelso, J. A. S. Are movements prepared in parts? Not under compatible (naturalized) conditions. *Journal of Experimental Psychology: General*, 1980, *109*, 475-495.

Goodwin, B. C. Biological stability. In C. H. Waddington (Ed.), *Towards a theoretical biology*. Chicago: Aldine, 1970.

Greene, P. H. Problems of organization of motor systems. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (Vol. 2). New York: Academic Press, 1972.

Grillner, S. How detailed is the central pattern generator for locomotion? *Brain Research*, 1975, *88*, 367-371.

Hallett, M., Shahani, B. T., & Young, R. R. Analysis of stereotyped voluntary movements at the elbow in patients with Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, 1977, *40*, 1129-1135.

Halliday, A. M., & Redfern, J. W. T. Finger tremor in tabetic patients and its bearing on the mechanism producing the rhythm of physiological tremor. *Journal of Neurology, Neurosurgery, and Psychiatry*, 1958, *21*, 101-108.

Hollerbach, J. M. *A study of human motor control through analysis and synthesis of handwriting*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, August, 1978.

Holst, E. von. Relative coordination as a phenomenon and as a method of analysis of central nervous functions. In *The behavioral physiology of animals and man. Selected papers of Eric von Holst*. Coral Gables, Fl.: University of Miami Press, 1973.

Horsley, V., & Schaffer, H. J. Experiments on the character of the muscular contractions which are evoked by excitation of the various parts of the motor tract. *Journal of Physiology*, 1886, *7*, 96-110.

Iberall, A. S. *Toward a general science of viable systems*. New York: McGraw-Hill, 1972.

Iberall, A. S. A field and circuit thermodynamics for integrative physiology: I. Introduction to general notions. *American Journal of Physiology/Regulatory, Integrative, & Comparative Physiology*, 1977, *2*, R171-R180.

Iberall, A. S. A field and circuit thermodynamics for integrative physiology: II. Keeping the books - a general experimental method. *American Journal of Physiology/Regulatory, Integrative, & Comparative Physiology*, 1978, *3*, R85-R97.

Kamen, G. *Fractionated reaction and reflex time after fatiguing isometric and isotonic exercise*. Unpublished doctoral dissertation, University of Massachussets, Amherst, 1980.

Katchalsky, S. K., Rowland, V., & Blumenthal, R. Dynamic patterns of brain cell assemblies. *Neurosciences Research Program Bulletin*, 1974, *12*.

Kelso, J. A. S. Contrasting perspectives on order and regulation in movement. In A. Baddeley & J. Long (Eds.), *Attention and performance, IX*. Hillsdale, N.J.: Erlbaum, 1981.

Kelso, J. A. S., & Holt, K. G. Exploring a vibratory system analysis of human movement production. *Journal of Neurophysiology*, 1980, *43*, 1183-1196.

Kelso, J. A. S., Holt, K. G., Kugler, P. N., & Turvey, M. T. On the concept of coordinative structures as dissipative structures: II. Empirical lines of convergence. In G. E. Stelmach (Ed.), *Tutorials in motor behavior*. New York: North-Holland Publishing Co., 1980.

Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. Patterns of human interlimb coordination emerge from the properties of non-linear, limit cycle oscillatory processes: Theory and data. *Journal of Motor Behavior*, 1981, *13*, 226-261.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North-Holland Publishing Company, 1980.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. On coordination and control in naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), *The development of movement control and coordination*. Chichester: John Wiley, 1982.

Marsden, C. D. The mechanisms of physiological tremor and their significance for pathological tremors. In J. E. Desmedt (Ed.), *Physiological tremor, pathological tremor and clonus: Progress in clinical neurophysiology* (Vol. 5). New York: Karger, 1978.

Matthews, P. B. C., & Muir, R. B. Comparison of electromyogram spectra with force spectra during human elbow tremor. *Journal of Physiology*, 1980, *302*, 427-441.

Morowitz, H. J. *Energy flow in biology*. Woodbridge, Conn.: Oxbow Press, 1979.

Prigogine, I. *From being to becoming*. San Francisco: W. Freeman, 1980.

Rack, P. M. H. Mechanical and reflex factors in human tremor. In J. E. Desmedt (Ed.), *Physiological tremor, pathological tremor and clonus: Progress in clinical neurophysiology* (Vol. 5). New York: Karger, 1978.

Rack, P. M. H., & Westbury, D. R. The short range stiffness of active mammalian muscle and its effect on mechanical properties. *Journal of Physiology*, 1974, *240*, 331-350.

Randall, J. E. A stochastic time series model for hand tremor. *Journal of Applied Physiology*, 1973, *34*.

Rietz, R. R., & Stiles, R. N. A viscoelastic-mass mechanism as a basis for normal postural tremor. *Journal of Applied Physiology*, 1974, *37*, 852-861.

Runeson, S. On the possibility of "smart" perceptual mechanisms. *Scandinavian Journal of Psychology*, 1977, *18*, 172-179.

Shik, M. L., & Orlovsky, G. N. Coordination of the limbs during running of the dog. *Biophysics*, 1965, *10*, 1148-1159.

Shik, M. L., & Orlovsky, G. N. Neurophysiology of locomotor automatism. *Physiological Reviews*, 1976, *56*, 465-501.

Sollberger, A. *Biological rhythm research*. Amsterdam: Elsevier, 1965.

Soodak, H., & Iberall, A. S. Homeokinetics: A physical science for complex systems. *Science*, 1978, *201*, 579-582.

Stein, P. S. G. Mechanisms of interlimb phase control. In R. H. Herman, S. Grillner, P. S. G. Stein, & D. G. Stuart (Eds.), Advances in behavioral biology, Vol. 18: Neural control of locomotion. 1976, 265-292.

Stein, R. B., & Lee, R. G. Tremor and clonus. In V. B. Brooks (Ed.), Handbook of physiology, Vol. 3: Motor control. Bethesda, Md.: American Physiological Society, 1981.

Stiles, R. N., & Randall, J. E. Mechanical factors in human tremor frequency. Journal of Applied Physiology, 1967, 23, 324-330.

Sutton, G. C., & Sykes, K. The variation in hand tremor with force in healthy subjects. Journal of Physiology, 1967, 191, 669-711.

Travis, C. E. The relation of voluntary movement to tremors. Journal of Experimental Psychology, 1929, 515-524.

Weiss, A. D. The locus of reaction time change with set, motivation, and age. Journal of Gerontology, 1965, 20, 60-64.

Yates, F. E. Physical biology: A basis for modeling living systems. Journal of Cybernetics and Information Science, 1979, 2, 57-70.

Yates, F. E., Marsh, D. J., & Iberall, A. S. Integration of the whole organism: A foundation for a theoretical biology. In J. A. Behnke (Ed.), Challenging biological problems: Directions towards their solution. New York: Oxford University Press, 1972, 110-132.

FOOTNOTE

[1] It is important to note that the four theoretical models do not generally have solutions in closed form. Thus, numerical integration was used to evaluate the probability density functions. By taking discrete time slices from the density function it was possible to determine the number of movement initiations expected within any particular phase of the tremor cycle. The resultant distributions could then be compared directly with the data obtained from each of the experiments.

APPENDIX 1

Derivation of the Models

All four models assume that tremor peak-to-peak time is distributed normally with some mean, μ_x, and variance, σ_x^2. Thus, if x is a random normal variable (r.n.v.) of tremor peak-to-peak time, that is distributed normally with some mean, μ_x, and variance, σ_x^2, then:

$$x \sim n(\mu_x, \sigma_x^2)$$

and the distribution function of x, f(x) is described as:

$$f(x) = \frac{1}{\sqrt{2\pi}\, S_x} e^{-1/2} \quad \frac{(x-\bar{x})^2}{S_x^2}$$

where $\bar{x}$ is the sample mean; S_x^2 is the sample variance.

The rationale and test of this assumption are given in Goodman (1981).

Model 1. Since y is distributed uniformly over the peak-to-peak interval x, then the distribution function of y given x, g(y|x) is described as:

$$g(y|x) = \frac{1}{X} I_{[o,x]}(y)$$

where y is a random variable defined as peak-to-movement initiation time.

Hence the conjoint distribution of peak-to-peak times and peak-to-movement initiation times is distributed as:

$$h(y,x) = \left[\frac{1}{X} I_{[o,x]}(y)\right] \left[\frac{1}{\sqrt{2\pi}\, S_x} e^{-1/2} \quad \frac{(x-\bar{x})^2}{S_x^2}\right]$$

After integrating over the limits of x, the resultant probability density function of peak-to-movement initiation time, y, is that given in equation (1) of text.

Model 2. A similar argument follows for model 2, which assumes a uniform distribution of y in the ascending phase of the peak-to-peak interval x. g(y|x) is described as:

$$g(y|x) = \frac{2}{X} I_{[\frac{1}{2} x, x]}(y)$$

Hence the conjoint distribution of tremor peak-to-peak times and peak-to-movement initiation times is distributed as:

$$h(y,x) = \left[\frac{2}{X} I_{[\frac{1}{2} x, x]}(y)\right] \left[\frac{1}{\sqrt{2\pi}\, S_x^2} e^{-1/2} \quad \frac{(x-\bar{x})^2}{S_x^2}\right]$$

By integrating over the limits of x, the probability density function of peak-to-movement time, y, given in equation (2) of text results.

Model 3. In model 3, y is a random normal variable distributed about x/2. g(y|x) is then described as:

$$g(y|x) = \frac{1}{\sqrt{2\pi}\ S_y} \quad e^{-1/2} \quad \frac{(y - \frac{\bar{x}}{2})^2}{S_y^2}$$

Hence the conjoint distribution of tremor peak-to-peak times and peak-to-movement initiation times is distributed as:

$$h(y,x) = \left[\frac{1}{\sqrt{2\pi}\ S_x} \quad e^{-1/2} \quad \frac{(x-\bar{x})^2}{S_x^2}\right] \left[\frac{1}{\sqrt{2\pi}\ S_y} \quad e^{-1/2} \quad \frac{(y - \frac{\bar{x}}{2})^2}{S_y^2}\right]$$

Thus, by integrating over the limits of x, the probability density function of peak-to-movement initiation time, y, given in equation (3) of text results.

Model 4. In model 4, y is a random normal variable distributed about 3x/4. g(y|x) is then described as:

$$g(y|x) = \frac{1}{\sqrt{2\pi}\ S_y} \quad e^{-1/2} \quad \frac{(y - \frac{3}{4}\bar{x})^2}{S_y^2}$$

Hence the conjoint distribution of tremor peak-to-peak times and peak-to-movement initiation times is distributed as:

$$h(y,x) = \left[\frac{1}{\sqrt{2\pi}\ S_x} \quad e^{-1/2} \quad \frac{(x-\bar{x})^2}{S_x^2}\right] \left[\frac{1}{\sqrt{2\pi}\ S_y} \quad e^{-1/2} \quad \frac{(y - \frac{3}{4}\bar{x})^2}{S_y^2}\right]$$

By integrating over the limits of x, the probability density function of peak-to-movement initiation time, y, given in equation (4) of text results. Thus the resultant probability density function of (4).

DIFFERENCES BETWEEN EXPERIENCED AND INEXPERIENCED LISTENERS TO DEAF SPEECH

Nancy S. McGarr+

Abstract. The study examines differences between experienced and inexperienced listeners in understanding the speech of the deaf. Listeners heard test words in three conditions: sentences, isolated, and segmented (the last being words produced in sentences, excised, and then presented in isolation). Factors believed influential in listener differences were examined: predicted word intelligibility, sentence context, sentence length, and position of the word in the sentence. Scores for experienced listeners were consistently higher than those for inexperienced listeners for all factors considered. Differences between listeners were greatest for test words in sentences, followed by isolated and segmented test words. However, there was no statistically significant interaction between listener experience and any of the factors considered. Thus, the data do not support several hypotheses that have been proposed to account for listener differences. For both experienced and inexperienced listeners, scores varied systematically depending on the amount of linguistic context in the sentence. In addition, a significant difference in scores for isolated and segmented test words suggests coarticulatory effects in the speech of the deaf that may significantly affect intelligibility for both groups.

INTRODUCTION

Those who work with the deaf are not suprised when a child whose speech is judged relatively intelligible in the classroom is still virtually unintelligible to the "man on the street." That there are judgment differences between experienced listeners (e.g., teachers of the deaf) and inexperienced listeners is widely accepted. In fact, intelligibility of deaf speech has been rated according to how likely the speaker is to be understood by "most trained teachers of the deaf, most people familiar with deaf speech, or almost everyone" (Thomas, 1963). In spite of this common observation, while consid-

---

+Also Center for Research in Speech and Hearing Sciences, Graduate School and University Center, The City University of New York.

Acknowledgment. This research was conducted in partial fulfillment of the requirements for the Ph.D. at the City University of New York. The author wishes to acknowledge the invaluable assistance of Harry Levitt and Katherine S. Harris who served as advisors to this work. Preparation of the manuscript was supported by Grants NS-13617 and 13870 to Haskins Laboratories.

erable effort has been directed to studying speaker characteristics for intelligibility, relatively little attention has been accorded factors related to listeners.

Investigators (Brannon, 1964; Markides, 1970; Smith, 1972) have noted that a naive listener may understand about one word in every five produced by a deaf speaker. In contrast, an experienced listener's ability to understand deaf speech seems clearly superior (Mangan, 1961; Markides, 1970; Monsen, 1978; Thomas, 1963). These studies used listeners to rate overall intelligibility or to transcribe speech production. Several differences between listening groups have been noted. First, intelligibility scores decreased from experienced to naive listeners (Mangan, 1961; Monsen, 1978; Nickerson, 1973; Thomas, 1963). Some overlap in individual data was observed, but as a whole, group scores for naive listeners never approached those of the experienced. For both groups, scores were higher for sentences than for isolated words with a wider range of intelligibility observed for sentences than for words (Hudgins, 1949; Subtelny, 1977; Thomas, 1963). Sentence scores for experienced listeners have been reported from 31% (Markides, 1970) to 83% (Monsen, 1978); sentence scores for inexperienced listeners ranged from 18.7% (Smith, 1972) to 73% (Monsen, 1978).

These data educe several hypotheses about listener differences. For example, the consistency of the reported speech production errors suggested to Hudgins and Numbers (1942) that the experienced listener may recode deaf speech to compensate for typical deaf articulatory errors. Since these error patterns are presumably unknown to the naive listener, articulatory cues cannot be used to enhance intelligibility. Hudgins and Numbers (1942) also hypothesized that experienced listeners may make better use of contextual information. They argued that the naive listener was so distracted by the quality of deaf speech that information could not be derived from available contextual cues. On the other hand, higher scores for sentences than for isolated words led Brannon (1964) to conclude that context was extremely important for the naive listener. Thomas (1963) noted that both groups profited from context, since scores for "everyday" sentences were higher than for isolated words. In these investigations, and others (Hudgins, 1949; Subtelny, 1977), context was defined as a word produced and heard in a sentence. However, the sentences varied considerably in the amount of linguistic information and different vocabulary was used in the sentence and isolated word conditions. Furthermore, for non-deaf speakers words produced in sentences differ from those produced in isolation (Lieberman, 1963; McGarr, 1981; Miller, Heise, & Lichten, 1951; O'Neill, 1957; Pollack & Pickett, 1963, 1964), although this difference has not been studied in deaf speakers.

Finally, in these studies, the criterion of listener experience was not always carefully controlled. In some instances experienced listeners were very familiar with the children, the speech training protocol, or the test material. In other studies, the listeners were not familiar with any of these factors. Many feel that it is personal knowledge of a particular deaf speaker that gives the experienced listener his or her advantage. But the extent to which each of these factors increases intelligibility of deaf speech for listeners has not been determined. This study was undertaken, therefore, to study systematically those factors believed to account for some of the differences between experienced and inexperienced listeners to deaf speech.

METHODS

Listeners

One hundred and twenty listeners participated in the study--sixty experienced and sixty inexperienced. An experienced listener was a person who had more than one year's experience in listening to the speech of the deaf. The sixty experienced listeners were teachers of the deaf, speech pathologists, and audiologists in schools for the deaf. The listeners did not know the child whose speech they heard or the school at which the child received training. The number of years of experience ranged from just over 1 year to 25 years; mean number of years' experience was 6.8 years. In addition to meeting the experience criterion, each of the listeners had normal hearing and was a native speaker of English.

An inexperienced listener was defined as having no previous experience in hearing the speech of the deaf. There were 60 inexperienced listeners recruited primarily from undergraduate classes. These listeners also met all other criteria required of the experienced group.

Subjects

Twenty severe-profoundly deaf children from the Lexington School for the Deaf served as subjects in the study. The children were equally divided into two age groups, one of 8- to 10-year-olds and another of 13- to 15-year-olds, with 5 females and 5 males in each group. All subjects were congenitally deaf and had no handicaps other than deafness. The group mean pure tone average for .5, 1, and 2 kHz was 98.6dB (ISO) in the better ear. The children were judged by their speech supervisors to have fair, average, or good speech. No child whose speech was judged totally unintelligible was included in the study.

Materials

The test materials comprised 36 monosyllabic words each of which was embedded in a sentence. The words were selected in order to examine possible interactions between listener experience and articulatory cues. Each word was empirically defined with respect to its predicted intelligibility when produced by a deaf child. This measure was obtained by ranking all words produced by deaf children in Smith's (1972) study. The 18 monosyllabic words ranked highest for intelligibility and the 18 monosyllabic words ranked lowest for intelligibility formed the test corpus. Scores for test words in the present study were subsequently compared with those of Smith and showed the same clustering of high and low intelligibility scores.

In order to examine the effect between listener experience and context, each of the 36 words was embedded in a sentence that varied with respect to the amount of overall contextual information. A definition of high or low contextual information was made for each of the sentences using a standard word prediction technique. Twenty undergraduates (not listeners) were asked to "fill-in the blank" when presented with a written version of the sentence with the test word omitted. A sentence was defined as high in contextual information if 15 or more undergraduates completed it with the same word. A sentence was defined as low in contextual information if 15 or more undergraduates selected different words to complete the sentence.

The sentences were also designed with respect to other factors that were believed to be important to listeners: (1) the number of syllables in the sentence, and (2) the location of the test word in the sentence. The sentences were either 3, 5, or 7 syllables in length; the location of the test word in the sentence occurred either (1) at or near the beginning of the sentence, (2) in the middle of the sentence, or (3) near or at the end of the sentence. Figure 1 is a schematic diagram summarizing key factors in the test materials. For the 36 test words in sentences, all factors in Figure 1 are relevant to the test material. For the test words in isolation, only predicted intelligibility is a factor. The test materials are presented in Appendix 1.

Listening Conditions

Since an isolated word differs from one in a sentence both in perception and production, an additional set of stimuli was produced maintaining the same balance of context and word intelligibility. Specifically these test words were originally produced in sentences but were subsequently heard by the listeners in isolation. These words are referred to as segmented test words and were obtained by processing the audio tape recordings of the childrens' sentences on the Haskins Laboratories spectrum and waveform editing system. Segmentation was accomplished using both auditory and visual cues. Because test words produced in sentences and isolation may vary in overall amplitude, the levels for the test words were equalized in each of the 3 listening conditions described below.

1. Test words produced in sentences and presented to the listener in sentences. Listeners were asked to write down the whole sentence; however, the scores for test words were of primary interest.

2. Test words produced in isolation and presented to the listener in isolation.

3. Test words produced in sentences, excised from the sentences, and presented to the listeners in isolation--segmented test words.

In each condition, the deaf speakers' samples were randomized in order to avoid learning effects. That is, each listener heard only one child with no repetition of the same test word on a tape. A single deaf child's intelligibility score was thus an average of 3 experienced and 3 inexperienced listeners' scores.

RESULTS

Intelligibility scores were obtained for experienced and inexperienced listeners, and analyses of variance performed to test for significant interactions between listener experience and other factors. Separate analyses were performed for test words in sentences, in isolation, and in segmented conditions because the number of factors was different for each type of

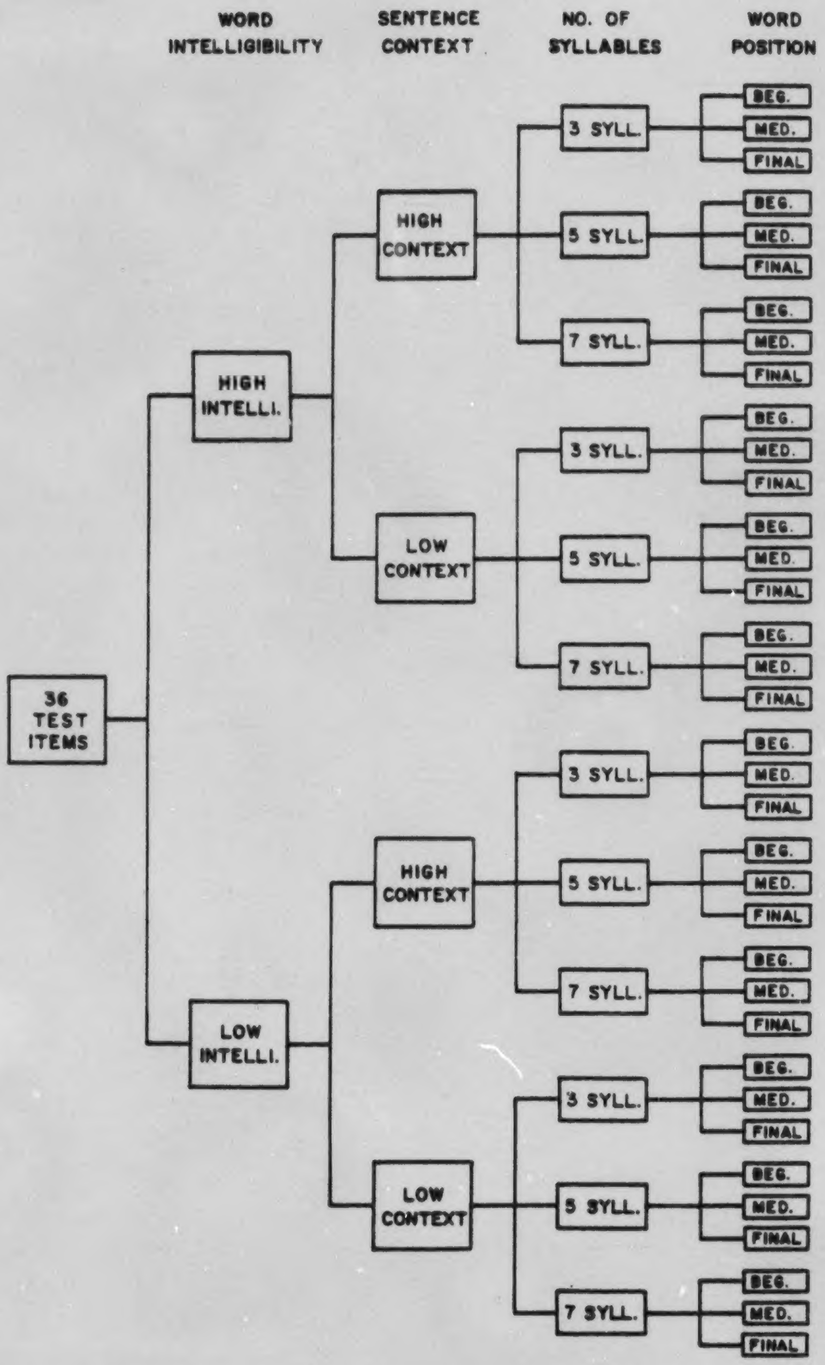

WORD INTELLIGIBILITY
SENTENCE CONTEXT
NO. OF SYLLABLES
WORD POSITION
36 TEST ITEMS
HIGH INTELLI.
LOW INTELLI.
HIGH CONTEXT
LOW CONTEXT
3 SYLL.
5 SYLL.
7 SYLL.
BEG.
MED.
FINAL

Figure 1. A schematic diagram summarizing the key factors in the test material. See text for further details.

stimulus. The factors considered in these analyses included listener experience, predicted word intelligibility, degree of sentence context, and two additional factors pertaining to the speakers: age of the children (younger versus older), and sex (male versus female). The analyses of variance for test words in sentences and for segmented test words included all five factors. The analysis for isolated words had only four factors since context was not a factor for words produced and heard in isolation.

In performing the analyses of variance, data were transformed using the arcsine transformation (Brownlee, 1965). Because of the large number of F tests performed in each of these analyses, only those effects with a significance level of .01 or smaller were considered. Table 1 summarizes data for each of the main effects as well as any significant interactions.

Listener experience was highly significant for test words in sentences and in isolation, but about the borderline significance level for segmented test words. There was no significant interaction between experience and any factor for test words in sentences or in isolation. There was evidence of a borderline interaction ($<.015$) between experience, intelligibility and context for segmented test words. Additional significant main effects included: context, predicted word intelligibility, and age (the latter factor was significant only for test words in sentences and in isolation). Sex was not a significant factor. There was evidence of an interaction between predicted word intelligibility and context (IxC) for test words in sentences.

In order to analyze the differences between the types of stimuli, a fourth analysis of variance was done. In this analysis the factors were: the type of stimulus (test words in sentences, in isolation, and segmented conditions), listener experience, and predicted word intelligibility. Each of the main effects was significant at the $< .01$ level. There were no significant interactions.

Listeners' Scores

Table 2 summarizes the mean scores obtained by experienced and inexperienced listeners for each type of speech stimulus. Experienced listeners consistently obtained higher scores than inexperienced listeners. For both groups, scores for test words in sentences were highest followed by scores for isolated words and then scores for segmented words. Scores for test words in sentences were more than double the scores for segmented words. The greatest difference between listeners occurred on sentences--11%. In contrast, the difference between listeners was 6% and 3% for words in isolation and for segmented test words, respectively. Intelligibility scores were also obtained for all words in sentences (cf. Table 2). Scores based on all words were only slightly higher than for scores based on test words alone.

Predicted Intelligibility of Test Words

Mean scores obtained by experienced and inexperienced listeners as a function of predicted intelligibility of test words are plotted in Figure 2. Experienced listeners obtained higher scores than inexperienced listeners for either high or low intelligibility words in sentence, in isolated, or in segmented conditions. The overall pattern of the data for high and low

Table 1

Source of Variation	Sum of Squares	DF	Mean Square	F	Significance
ANALYSIS OF VARIANCE FOR TEST WORDS PRODUCED AND HEARD IN SENTENCES					
Experience (E)	2.44	1	2.44	20.5	.001*
Context (C)	6.38	1	6.38	53.61	.001*
Word Intell. (I)	2.73	1	2.73	22.94	.001*
Age (A)	12.04	1	12.04	11.58	.003*
Sex (S)	1.07	1	1.07	1.03	.326
IxC	2.20	1	2.20	18.50	.001*
ANALYSIS OF VARIANCE FOR TEST WORDS PRODUCED AND HEARD IN ISOLATION					
Experience (E)	.442	1	.440	14.6	.001*
Word Intell. (I)	2.313	1	2.310	70.9	.001*
Age (A)	3.08	1	3.08	11.84	.003*
Sex (S)	.10	1	.10	.38	.542
ANALYSIS OF VARIANCE FOR TEST WORDS PRODUCED IN SENTENCES AND HEARD IN ISOLATION (SEGMENTED)					
Experience (E)	.292	1	.292	5.03	.025
Context (C)	1.740	1	1.740	30.00	.001*
Word Intell. (I)	.884	1	.884	15.24	.001*
Age (A)	.706	1	.706	4.46	.048
Sex (S)	1.206	1	1.206	7.63	.013**
ExIxC	.342	1	.342	5.89	.015**

*Significant at < .01 level
**Significant between .01 and .02 levels

Table 2

Mean Scores Obtained by Listeners

Type of Stimulus	Listeners	Mean Score % Correct
Test words produced and heard in sentences	Experienced	.41
	Inexperienced	.30
Test words produced and heard in isolation	Experienced	.29
	Inexperienced	.23
Test words produced in sentences and heard in isolation (i.e. segmented)	Experienced	.16
	Inexperienced	.13
All words produced and heard in sentences	Experienced	.49
	Inexperienced	.35

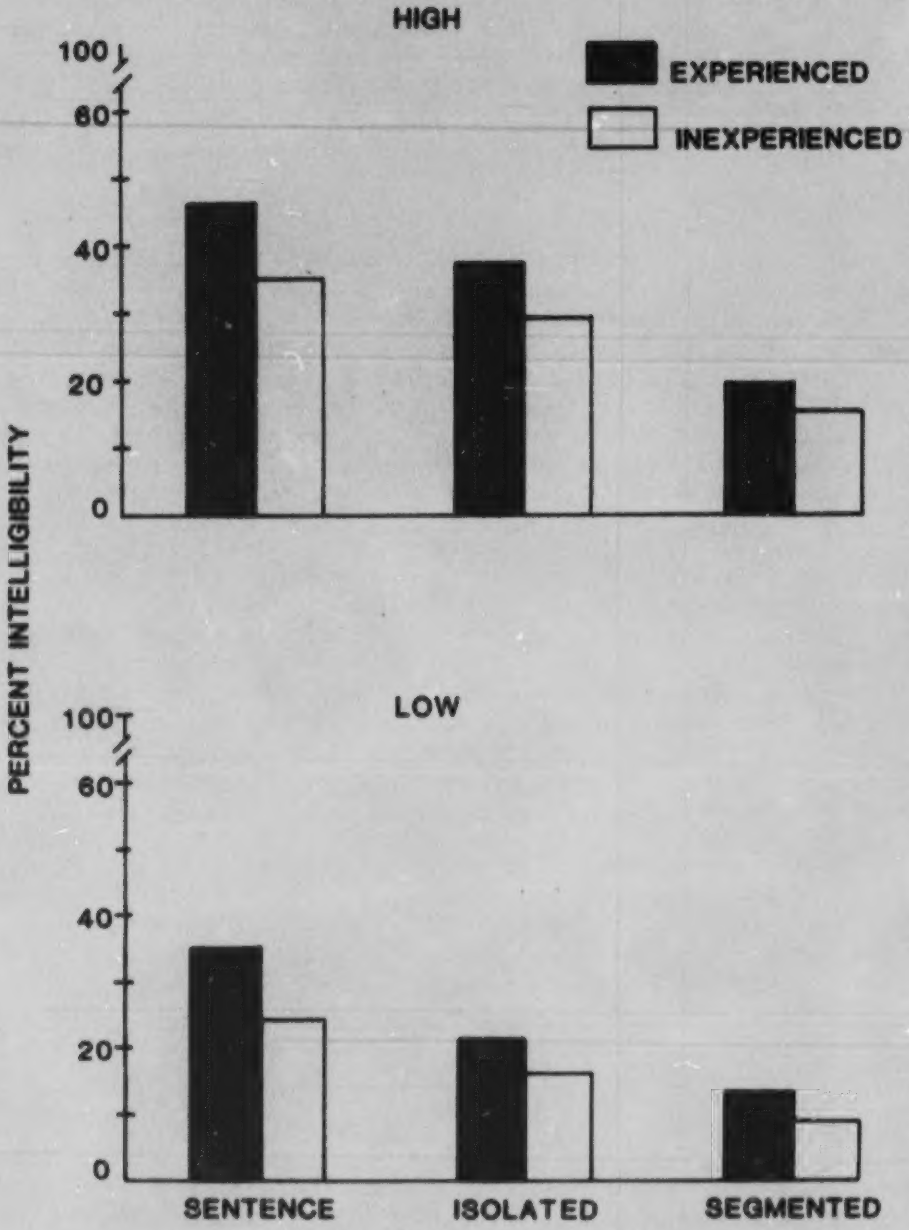


Figure 2. Mean scores obtained by experienced and inexperienced listeners for test words in sentences, in isolated and in segmented conditions. Data are graphed as a function of predicted word intelligibility (high or low).

intelligibility words was similar for both groups. Test words with high predicted intelligibility received higher scores than those with low predicted intelligibility for each type of stimulus. For either high or low intelligibility words, scores were highest when the test words were in sentences, followed by test words in isolation, and finally segmented test words. However, the effect of intelligibility was most pronounced for test words in sentences and in isolation. In these conditions, scores obtained by both groups of listeners were noticeably higher for test words with high predicted intelligibility than with low. High or low intelligibility had less effect on the scores for segmented words. There was no statistically significant interaction between intelligibility and stimulus type.

Sentence Context

Mean scores obtained by experienced and inexperienced listeners for test words as a function of sentence context are plotted in Figure 3. For all conditions, experienced listeners scored higher on average than inexperienced listeners but again, no statistically significant interaction was found. The differences between experienced and inexperienced listeners for test words in either high or low context sentences was roughly 10%. Since segmented test words were originally produced in sentences, the effect of context on intelligibility of these stimuli was also examined. The difference between listeners for segmented words produced in high or low context sentences was roughly 5%.

The magnitude of the context effect is also evident in Figure 3. Scores for both groups of listeners were greater for the high context conditions than for the low. Scores for test words in high context sentences were approximately 16% greater than those in low context sentences for listeners. For segmented test words, difference between high and low context conditions was approximately 8% for either group. Thus, the effect of context for words produced and heard in sentences is substantial. If the same test words are segmented in such a way that, although produced in context they are heard in isolation, the effect of context is much smaller, but not negligible.

Interaction Between Experience, Context, and Intelligibility

Of special interest was the significant interaction between intelligibility and context for sentences as well as any interaction involving experience and these factors. The interactions between context and predicted intelligibility (IC) were statistically significant for test words in sentences. A borderline interaction was obtained for listener experience, context and predicted word intelligiblity (EIC) for segmented test words. These three factors are plotted in Figure 4.

For test words in sentences, the pattern for experienced and inexperienced listeners is similar, with the difference between listeners averaging about 10% across each of the four combinations of intelligibility and context. For both groups of listeners, the ranking of scores (from highest to lowest) as a function of predicted intelligibility and sentence context were: (1) high intelligibility, high context, (2) low intelligibility, high context, (3) high intelligibility, low context, and (4) low intelligibility, low context.

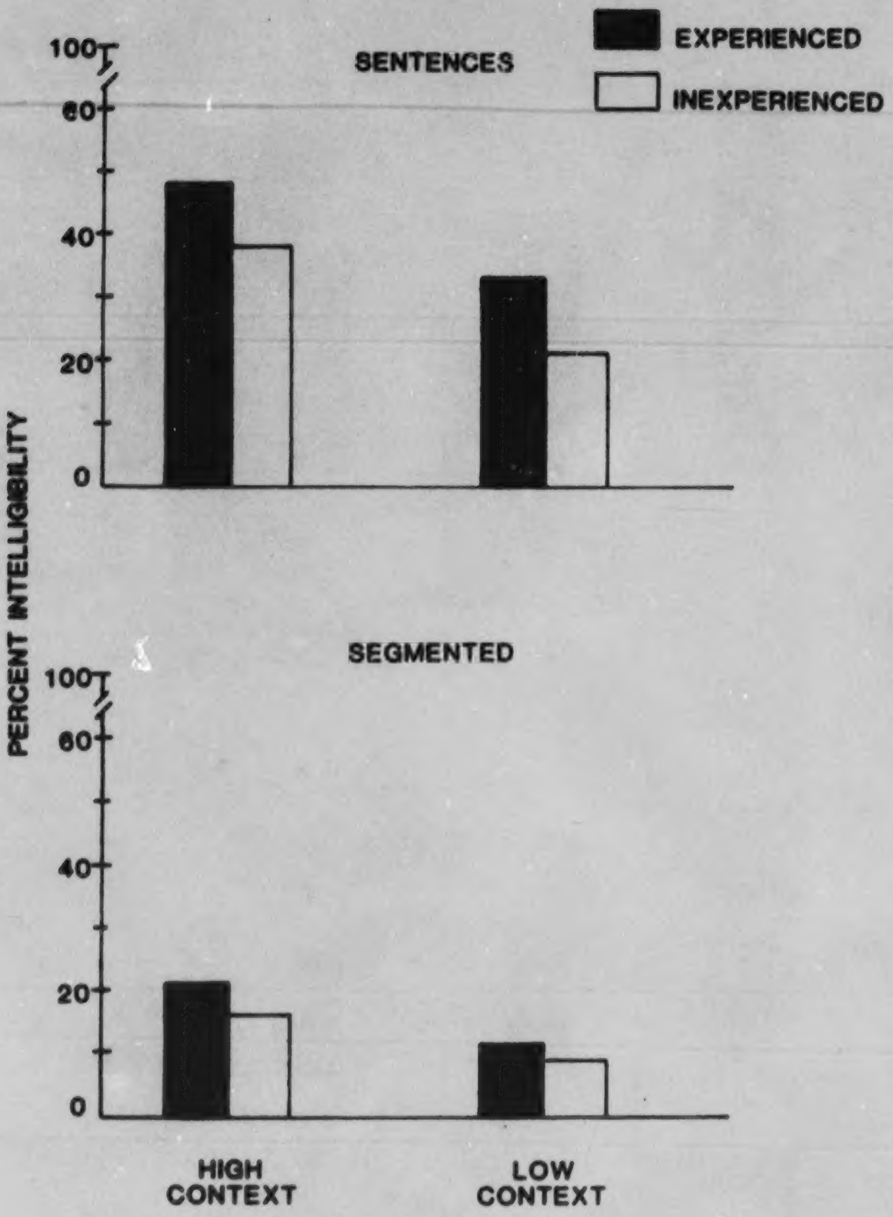


Figure 3. Mean scores obtained by experienced and inexperienced listeners for test words graphed as a function of high or low context.

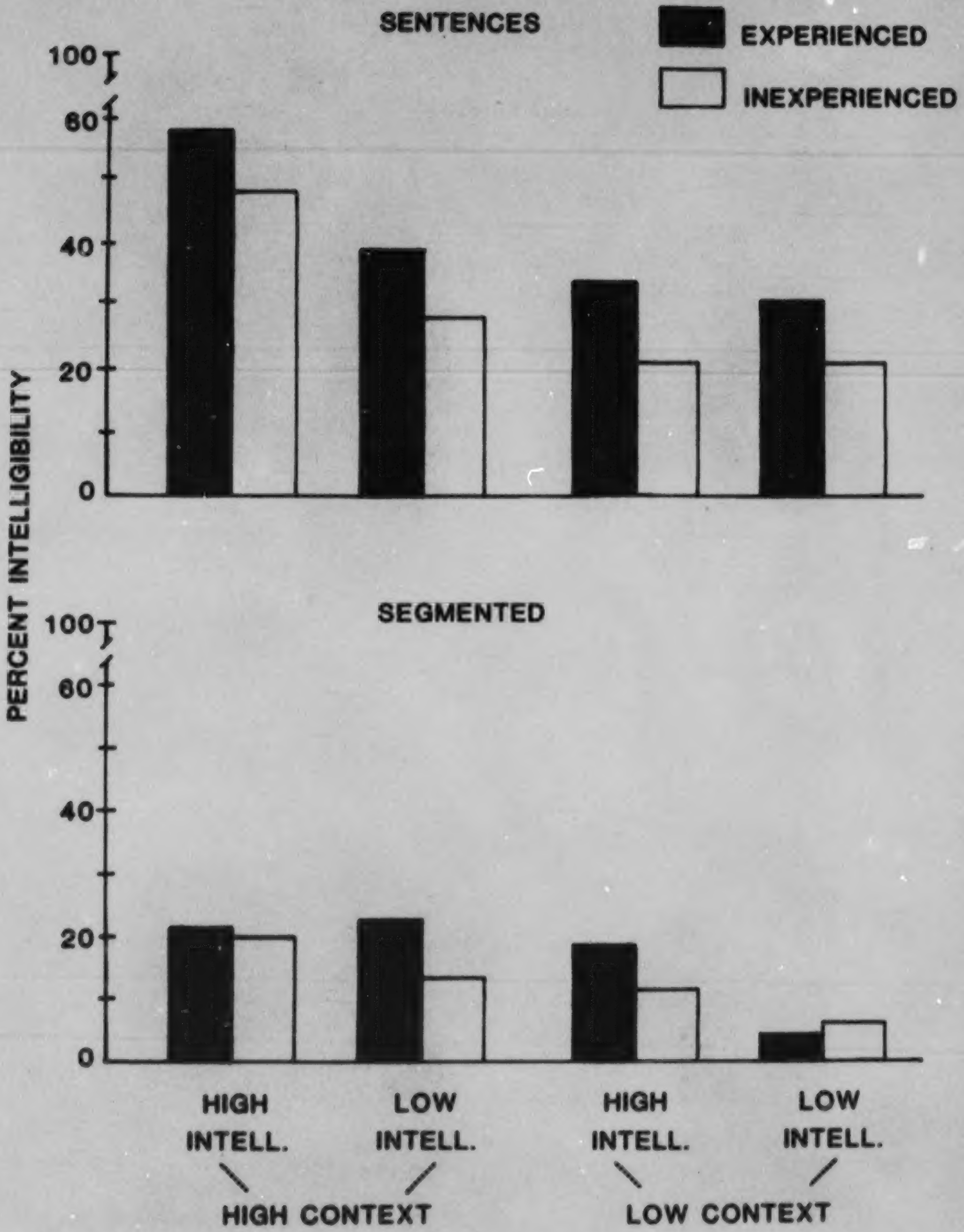


Figure 4. Mean scores obtained by experienced and inexperienced listeners for test words plotted as a function of predicted word intelligibility and context.

For segmented test words, the overall patterns for experienced and inexperienced listeners show relatively the same ranking of intelligibility as for the sentence condition. That is, for both experienced and inexperienced listeners, high context words were most intelligible and low context words, least intelligible. Also, on average, scores for test words with high intelligibility were higher than those with low intelligibility. In only one instance did inexperienced listeners receive slightly higher scores than experienced listeners. That is, for segmented test words with low context, the experienced listeners showed a significant drop in scores from high to low intelligibility words. This gives rise to the borderline interaction.

Between Children Differences

Intelligibility scores were also analyzed for factors related to the children's age and sex. These data are shown in Figure 5. Again, there were no interactions between listener experience and these variables. As indicated by the analysis of variance, age was a significant factor for test words in sentences and in isolation, but not for segmented test words. Older children were more intelligible than younger children for all three types of stimuli. Further, there were no significant differences between male and female subjects for test words in sentences and isolation, and only a borderline significance level for segmented test words.

Position of the Test Word and Number of Syllables

An additional analysis of variance was performed to investigate the effect of the position of the test word in the sentence, the number of syllables in the sentence, and whether there were any interactions between listener experience and these two factors.

The main effect for position of the test word in the sentence was highly significant ($p < .001$). No statistically significant effect was found for the number of syllables in the sentence. However, there was a statistically significant interaction ($p < .001$) between the number of syllables in the sentence and the position of the word in the sentence. Again, there was no statistically significant interaction between listener experience and these factors.

Figure 6 shows the percent intelligibility obtained by listeners for test words as a function of position in the sentence. Again experienced listeners obtained higher scores than the inexperienced listeners regardless of word position. For test words in sentences, the pattern of relative intelligibility was similar for both groups. Scores were highest for test words near the beginning of sentences, followed by those in the middle, and those near the end of sentences. In the sentence condition, the difference between experienced and inexperienced listeners was approximately 10% for each position. In contrast, experienced listeners scored only slightly higher than inexperienced for segmented test words. The difference between groups was only 5%; scores for test words segmented from the beginning, middle, or end of the sentences were nearly the same. There was no significant interaction between listener experiences and position of the test word.

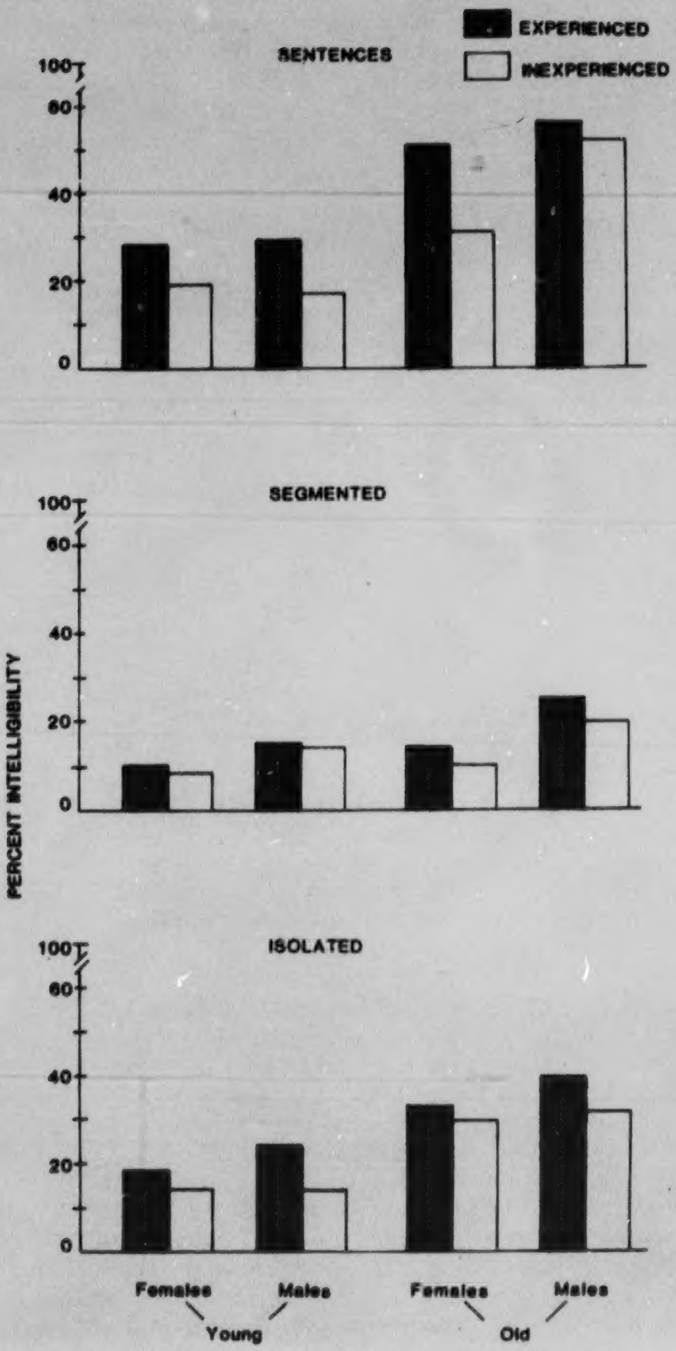


Figure 5. Mean scores obtained by experienced and inexperienced listeners for test words plotted as a function of the subjects' age and sex.

48

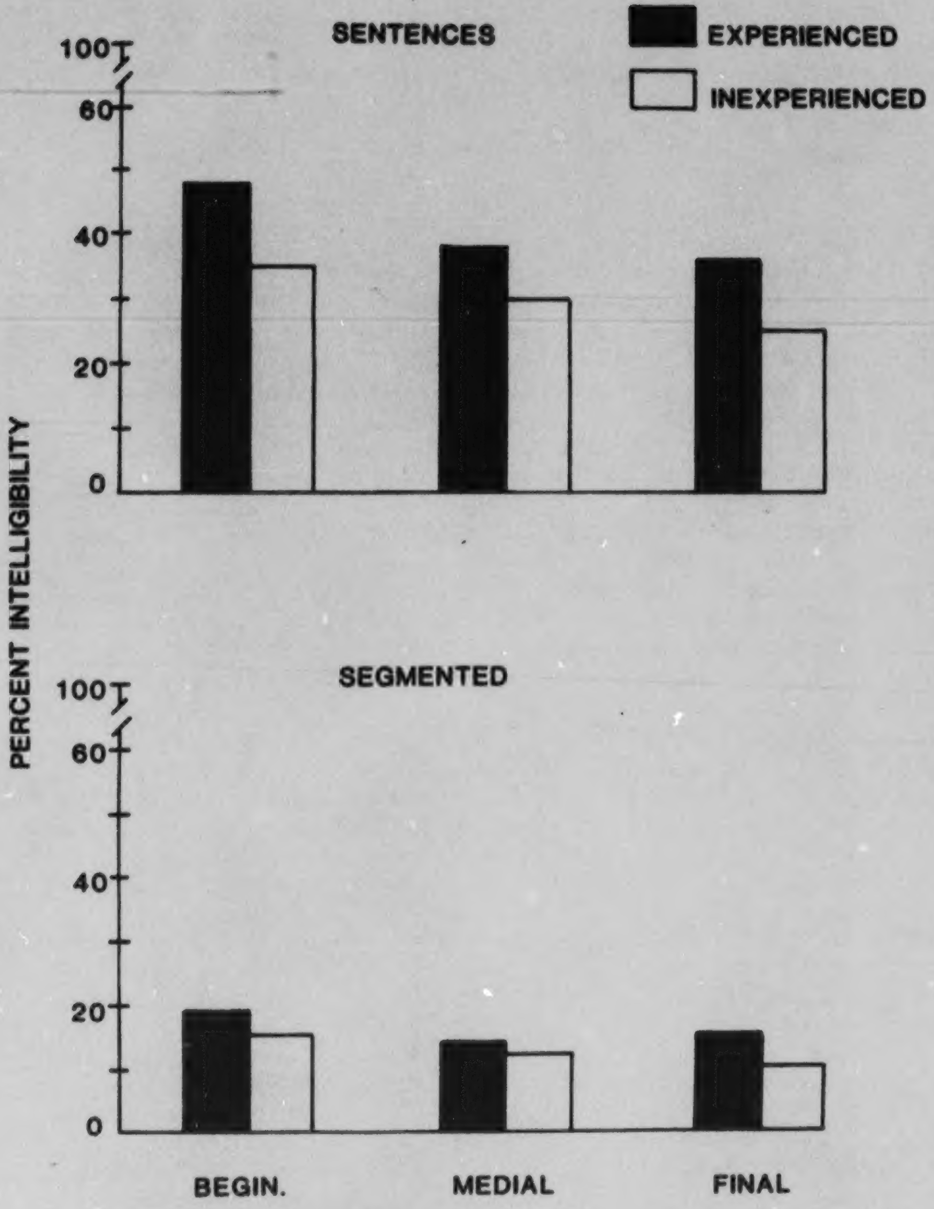


Figure 6. Mean scores obtained by listeners as a function of the position of the test word in the sentence.

Figure 7 plots the significant interaction between number of syllables and word position in the sentence for both groups of listeners. There was no interaction effect for test words in the segmented condition. For three-syllable sentences, test words at the beginning of the sentence were less intelligible than those near the beginning of five- and seven-syllable sentences. It should be noted that the test words in three-syllable sentences were always in the word initial position, while those in the five- and seven-syllable sentences occurred near (within two syllables) the beginning of the sentence but not in the word initial position. Differences between experienced and inexperienced listeners were greatest for test words near the beginning of five-syllable sentences, and for test words near the middle and end of seven-syllable sentences.

DISCUSSION

Intelligibility scores for the experienced listeners were consistently higher than those for inexperienced listeners. Further, the differences in the test scores between experienced and inexperienced listeners were essentially constant for all factors investigated: (1) predicted word intelligibility, (2) degree of sentence context, (3) number of syllables in the sentence, and (4) position of the test word in the sentence. For both groups of listeners, the scores for test words in sentences were consistently higher than scores for test words in isolation followed by segmented words.

Where comparisons are possible, these data are not inconsistent with the literature. For words produced and heard in isolation, the scores obtained by experienced listeners are reported from 35% (Subtelny, 1977) to 42% (Hudgins, 1949); the mean score for experienced listeners in this study was 29%. For inexperienced listeners, the reported scores range from 17% (Brannon, 1964) to 28% (Thomas, 1963); mean score obtained by the inexperienced listeners in this study was 23%. Test words with high predicted intelligibility fell essentially mid-range of the published data for either experienced or inexperienced listeners. This suggests that phonetically balanced monosyllables frequently chosen as the speech stimuli for deaf subjects are similar to test words with high predicted intelligibility used in this study. Choice of phonetically balanced monosyllables in speech evaluations would likely result in higher intelligibility scores for deaf speakers than if other word lists were chosen.

Scores reported for sentences vary over a wider range of intelligibility than those for isolated words. For experienced listeners, scores are reported from 31% (Markides, 1970) to 83% (Monsen, 1978); for inexperienced listeners, the range was 18.7% (Smith, 1972) to 73% (Monsen, 1978). Scores for test words in sentences in this study were 41% for experienced, and 30% for inexperienced listeners, with scores for all words in sentences only slightly higher (49% and 35%, respectively).

If sentence scores from this study are examined as a function of context, the scores for high context sentences were 49% for experienced and 38% for inexperienced listeners and nearly mid-range of data reported in the literature. Scores for sentences with low context were 33% for experienced, and 21% for inexperienced listeners and fell near the lower end of the reported range for the respective groups. Apart from the present study, which controlled for

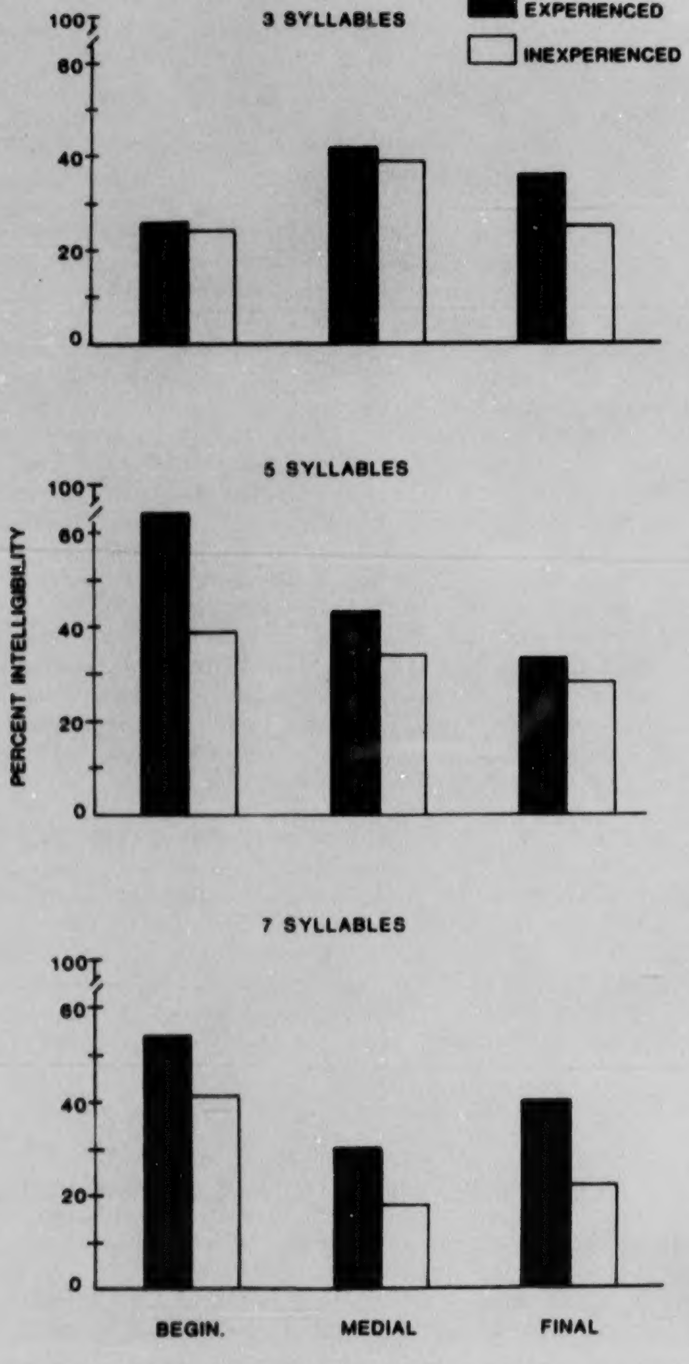


Figure 7. Mean scores obtained by listeners as a function of the position of the test word in the sentence and the number of syllables in the sentence. Data are for test words in the sentence condition.

the degree of context, the speech materials resulting in high intelligibility were those that contained words of common usage or were highly redundant in linguistic information. (e.g., Thomas, 1963; Monsen, 1978). Speech materials that resulted in lower intelligibility scores were either spontaneous speech samples (John & Howarth, 1965; Markides, 1970) or sentences that varied considerably in length and grammatical complexity (Smith, 1972). This wide variation in intelligibility scores reported for deaf children with very similar hearing losses implies the necessity for a set of uniform speech materials, thus permitting more meaningful evaluation of intelligibility, and also better comparison among deaf speakers.

These data do not, however, support several hypotheses that have attempted to explain the differences between listeners. Hudgins and Numbers (1942) proposed that experienced listeners obtained higher scores than inexperienced listeners because they are familiar with typical errors in production of deaf speech, and recode the speech so as to compensate for these errors. If this were the case, one would expect an interaction between listener experience and predicted word intelligibility. By definition, words with high intelligibility were ones that deaf children were likely to produce correctly. Similarly, words with low intelligibility were ones that deaf children were likely to misarticulate. Hence, if the above hypothesis was correct, experienced listeners would show a greater relative gain for low intelligibility words, since these words should have more errors for the listener to recode. However, no significant interaction was obtained. The measured difference in scores between experienced and inexperienced listeners for test words with high intelligibility was about the same as those for test words with low intelligibility, as shown in Figure 2. The lack of a statistically significant interaction between listener experience and predicted word intelligibility does not mean that experienced listeners recode deaf speech in the same way as inexperienced listeners, but rather that recoding strategies are more subtle and less easily defined than previously proposed.

A second hypothesis (Hudgins & Numbers, 1942; Thomas, 1963), proposes that experienced listeners simply make better use of contextual cues. Scores for both classes of listeners were higher for sentences with high context than for those with low context (cf. Figure 3) and there was no evidence of a statistically significant interaction between listener experience and context. The improvement due to experience was essentially constant for both high context and low context stimuli. Again, the lack of a statistically significant interaction does not repudiate the importance of context, but rather indicates that should an interaction exist, it is likely to be of a smaller magnitude than suggested.

While the effect of context on speech intelligibility has long been realized, it had been argued by Hudgins and Numbers (1942) that context may be even more important for listeners of deaf speech. Specifically, they hypothesized that the effect of articulatory errors on the intelligibility of deaf speech could be reduced by the contextual constraints of the sentences, and by implication, the greater the articulatory errors, the greater the effect of context. This third hypothesis concerning an interaction between intelligibility and context was supported by the data. The effect of word intelligibility, from high to low, accounted for a greater change in scores for high context sentences than for low context sentences (cf. Figure 4, top). While there was a significant interaction between intelligibility and context for test words in sentences, the interaction between these factors and listener experience was not statistically significant, suggesting that both experienced and inexperienced listeners are benefiting to the same extent from this information. This effect was observed even for individual children whose intelligibility scores were low (<30%) (cf. McGarr, 1978). These results contravene Sitler, Schiavetti, and Metz (in press) who found no effect of context for subjects with poor intelligibility. It should be noted that Sitler et al. did not control for the degree of context in their test materials and also used different vocabulary for their isolated words and sentences.

A fourth view is that personal knowledge of the deaf speaker which enables the experienced listener to obtain higher intelligibility scores. Since the inexperienced listener does not know the speaker, his or her scores would be lower. In the literature, a definition of experienced listener included persons who knew the subjects, such as teachers or parents (Mangan, 1961), listeners who were trained on either the test materials or the deaf speakers (Hudgins, 1949), as well as listeners who were generally familiar with the speech of the deaf, but did not personally know the speakers. In contrast, all inexperienced listeners were specified as having no previous experience with the deaf. In this investigation, none of the listeners, experienced or inexperienced, knew the child whose speech they heard. Hence, the hypothesis of personal knowledge of the speaker alone enabling the experienced listener to obtain higher intelligibility scores was not supported in the study (see also Gulian & Hinds, 1981). While it is likely that children who are known to parents or teachers may be more intelligible than to other listeners, further research is warranted to quantify the effect of personal knowledge.

A final notion is that knowledge of a particular speech teaching strategy results in a distinctive speech pattern, characteristic of the child's school, which enables the experienced listener who is cognizant of these strategies to obtain higher intelligibility scores. Similarly, if other experienced listeners, or inexperienced listeners, are unfamiliar with this educational approach, the intelligibility scores will be lower. This view is also not supported by the data. Although the error patterns of the subjects are not discussed in detail here (cf. however, McGarr, 1978), the error patterns were similar to other deaf children (Smith, 1972; Levitt et al., Note 1). Also, the experienced listeners in this study did not know at which school the child was trained. Teachers serving as experienced listeners who were from the same school as the children scored no better or worse than the experienced listeners from other schools. It would seem that once familiar with deaf speech, the experienced listeners were able to generate higher scores for deaf speakers in general.

One can infer from the results of this study that the effect of context is important in perception as well as in production. For the former, the effect of linguistic context was seen in the differences in test scores for speech stimuli with high or low context, and also in the differences between test words produced and heard in sentences, and test words produced in sentences but heard in isolation (i.e., segmented). It should be remembered that the recordings of test words in sentences and in segmented conditions were identical. These results are described in greater detail elsewhere (McGarr, 1981).

The effect of phonetic context on production is noted in the differences in test scores between isolated words and segmented test words, the scores for the former being considerably higher. The difference in test scores indicates that deaf children produce words in context differently than words in isolation. This finding has been observed for hearing speakers (Lieberman, 1963; McGarr, 1981; Miller, Heise, & Lichten, 1951; O'Neill, 1957; Pollack & Pickett, 1963, 1964) but heretofore has not been quantified for deaf speakers. The data in this study suggest that deaf speakers do not produce speech "like-beads-on-a string" (Haycock, 1933). Rather, coarticulation occurs in the speech of the deaf and significantly affects intelligibility. It would be wrong, however, to assume that, since this effect seems to be a negative one (manifested by relatively low scores for segmented test words), the deaf child should be taught to produce speech one-word-at-a-time in order to improve intelligibility. While this study did not consider test words produced in isolation but heard in context, it is well known that speech produced by the concatenation of isolated words, without additional processing (Flanigan, 1972), is both difficult to understand and unpleasant to hear.

Another production effect observed was that the total energy for a word produced in isolation was different from that for the same word produced in sentences. Specifically, isolated test words tended to be more intense than those produced in sentences, and longer in duration. However, the perceptual differences observed in the study between test words in sentences and in isolation cannot be ascribed to differences in intensity, since the levels for test words in each condition (sentences, isolation, and segmented) were equalized.

Of the variables considered in this study, only the stimulus type (test words in sentences, in isolation, or in segmented conditions) showed any evidence of a possible interaction with listener experience. That is, the difference between experienced and inexperienced listeners was greater in sentences than in isolation. The finding of no significant interaction between listener experience and any factor investigated implies that the effect of experience is not due to any superficial recoding of deaf speech on the part of the listener. If the factors considered in this study (i.e., context, predicted word intelligibility, sentence length, or word position) were the keys to the differences between listeners, then marked improvement in the intelligibility of deaf speech for the "man on the street" could be accomplished by a training program that concentrated on those factors most responsible for the differences between listeners.

In addition to the main effects tested, it is also known that the difference between experienced and inexperienced listeners was not due to any secondary effects such as idiosyncracies in particular children or in specific test words. Overall scores for younger children were slightly poorer than those for older children, as was also observed by Smith (1972), and there was little difference between male and female speakers. Similarly, examining the scores obtained by experienced and inexperienced listeners for individual test words did not reveal any unusual variation from the patterns obtained for any other variables in the study.

In sum, the difference between experienced and inexperienced listeners cannot be accounted for in any obvious way. For each factor, analysis of the data indicates a remarkably constant difference between groups. The result of this finding suggests that the advantage of experience cannot be attributed simply to one or two variables, at least for the factors considered within this study. Consequently, the differences between experienced and inexperienced listeners must be due to fairly complex aspects of deaf speech that are not immediately apparent to the listener, but that must be learned. The fact that the difference between listeners was constant suggests that the effect occurs fairly consistently over a wide range of variables and there is a need for additional research. Such research might include studies of the effect of the personal knowledge of the speaker; the importance of visual cues; how spectral information in the speech of the deaf is coded differently from that of normals; and how coarticulatory phenomena are manifested in the speech of the deaf.

REFERENCE NOTE

1. Levitt. H., Stark, T., McGarr, N. S., Carp, J., Stromberg, H., Gaffney, R. S., Barry, C., Velez, A., Osberger, M. J., Leiter, E., & Freeman, L. *Language and communication skills of deaf children*. Unpublished report, 1973-1976.

REFERENCES

Brannon, J. B. *Visual feedback of glossal motions and its influence on the speech of deaf children*. Unpublished doctoral dissertation, Northwestern University, 1964.

Brownlee, K. A. *Statistical theory and methodology in science and engineering*. New York: John Wiley and Sons, 1965.

Flanigan, J. L. *Speech analysis synthesis and perception*. New York: Springer-Verlag, 1972.

Gulian, E., & Hinds, P. Response bias in the perception of deaf speech by experienced and naive listeners. *British Journal of Audiology*, 1981, *15*, 165-171

Haycock, G. S. *The teaching of speech*. Washington, D.C.: A. G. Bell Association, 1933.

Hudgins, C. V. A method of appraising the speech of the deaf. *Volta Review*, 1949, *51*, 597-638.

Hudgins, C. V., & Numbers, F. C. An investigation of the intelligibility of the speech of the deaf. *Genetic Psychology Monographs*, 1942, *25*, 289-392.

John, J. E. J., & Howarth, J. N. The effect of time distortions on the intelligibility of deaf children's speech. *Language and Speech*, 1965, *8*, 127-134

Lieberman, P. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 1963, *6*, 172-187.

Mangan, K. Speech improvement through articulation testing. *American Annals of the Deaf*, 1961, *106*, 391-396.

Markides, A. The speech of deaf and partially hearing children with special reference to factors affecting intelligibility. *British Journal of Disorders of Communication*, 1970, *5*, 126-140.

McGarr, N. S. *The differences between experienced and inexperienced listeners in understanding the speech of the deaf*. Unpublished doctoral dissertation, The City University of New York, 1978.

McGarr, N. S. The effect of context on the intelligibility of hearing and deaf children's speech. *Language and Speech*, 1981, *24*, 255-264.

Miller, G. A., Heise, G. A., & Lichten, W. The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*, 1951, *41*, 329-335.

Monsen, R. B. Toward measuring how well hearing impaired children speak. *Journal of Speech and Hearing Research*, 1978, *21*, 197-219.

Nickerson, R. S. *Computerized speech-training aids for the deaf* (Report No. 2366). Cambridge, Mass.: Bolt Beranek and Newman, Inc., 1973.

O'Neill, J. Recognition of intelligibility test materials in context and isolation. *Journal of Speech and Hearing Disorders*, 1957, *22*, 87-90.

Pollack, I., & Pickett, J. M. The intelligibility of excerpts from conversation. *Language and Speech*, 1963, *6*, 165-171.

Pollack, I., & Pickett, J. M. Intelligibility of excerpts from fluent speech: Auditory versus structural context. *Journal of Verbal Learning and Verbal Behavior*, 1964, *3*, 79-84.

Sitler, R., Schiavetti, N., & Metz, D. Contextual effects in the measurement of hearing impaired speakers' intelligibility. *Journal of Speech and Hearing Research*, in press.

Smith, C. R. *Residual hearing and speech production of deaf children*. Unpublished doctoral dissertation, City University of New York, 1972.

Subtelny, J. D. Assessment of speech with implication for training. In F. Bess (Ed.), *Childhood deafness*. New York: Grune and Stratton, 1977, 183-194.

Thomas, W. Intelligibility of the speech of deaf children. *Proceedings of the International Congress on the Education of the Deaf*. *Document No. 106*. Washington, D.C.: U. S. Government Printing Office, 1963.

Appendix 1

Test Sentences recorded by the deaf subjects.
The test word is underlined in each sentence.

High Context

3 Syllables

Keep quiet.
Read the book.
Come with me.
The dog barks.
Comb your hair.
That's no good.

5 Syllables

The cat chased the mouse.
My name is Nancy.
Get your coat and hat.
Get your ball and bat.
Did you brush your teeth?
Is there no more milk?

7 Syllables

That man is not my father.
I wish I had a pony.
We have food for the picnic.
The flag is red, white and blue.
May I have a piece of cake?
Can you dive in deep water?

Low Context

3 Syllables

Feed the dog.
Have a lot.
You did it.
I need it.
Get the cake.
This is his.

5 Syllables

They will come again.
Is that the tall one?
Mother has the car.
Who wants this ice cream?
It's easy to hear her.+
He said he could go.

7 Syllables

The book is on the table.
What was the name of that boy?
If it's cool I cannot go.
Is the fat baby crying?
It is nice on a fall day.
We will go to the beach today.+

+These sentences contain an additional syllable.

A LANGUAGE-ORIENTED VIEW OF READING AND ITS DISABILITIES*

Isabelle Y. Liberman+

For the past 15 years or so, my main research interest has been in early reading acquisition and the problems associated with it. During all that time, my colleagues and I in the Haskins Laboratories reading research group have been stressing the importance of language and the alphabet in the reading process, and, consequently, in its disabilities.

For most of that period, however, we (and a remarkably small number of other investigators) were rather lonely warriors battling against a massed field of special educators with quite different ideas about reading disabilities. Most numerous in the early years were the practitioners in schools, hospitals, and optometrists' offices, who approached the reading problem armed with balance beams, trampolines, parquetry blocks, strings of wooden beads, swinging balls suspended from the ceiling, and the like. The activities using this equipment were expected to improve the children's gross and fine motor coordination, which in turn were considered to be the foundation of visual perception, and then eventually were meant to correct deficits in visual perception itself, which were purported to be the root cause of reading problems.

Common sense had little place in all this. Simply ignored was such contrary evidence as the fact that spectacularly coordinated animals, including the great apes and some humans in professional athletics, had excellent visual perception but could not read, while their poorly coordinated, indeed, even crippled, brothers and sisters, whether seeing or nearly blind, might be fluent readers. Moreover, little research was directed toward actually exploring the verity of the hypothesis or the efficacy of the remediation based solely upon it (luckily for the children under their charge, many practitioners of this persuasion hedged their bets by adding daily reading remediation to their gymnastic and visual perceptual routines). When such

---

*To appear in H. Myklebust (Ed.), Progress in learning disabilities, Vol. 5. New York: Grune & Stratton.

+Also University of Connecticut.

Acknowledgment. Much of the work I have reported here reflects an interdisciplinary effort by various members of the reading research group of Haskins Laboratories. The group, which has included educators, psychologists, and linguists and their students, has worked together in harmony and with remarkable productivity over some 15 years. I wish here to acknowledge my debt to all of them, but would single out one member of the Laboratories, Alvin M. Liberman, for particular mention in gratitude for the many seed ideas and insights he has contributed so generously over the years since our reading research began.

[HASKINS LABORATORIES: Status Report on Speech Research SR-70 (1982)]

questions were at long last examined with care (Hammill, 1972), the evidence was found to be, indeed, strongly opposed to the view that poor motor coordination and visual perception were the root causes of most reading problems or that most reading problems could be eliminated by means of gymnastic and visual exercises. One could dare hope, then, that such procedures would finally be seen as useful for the remediation of other problems present in some poor readers, problems like clumsiness or poor visuomotor coordination, but not for reading remediation, and that such procedures would perhaps produce better ball players and bicyclists, but not necessarily better readers.

Recently, the situation did appear to be improving. There was more emphasis on language development and language processing in the special education journals. The teachers in the field were beginning to question the old routines; the teacher-trainers and the new special education texts seemed to be increasingly language-oriented. Publishers began putting "linguistic" in the titles of their reading series for the elementary grades and in the brochures used to promote their offerings--"linguistic" had clearly become a buzz word for "a good thing."

Unfortunately, it appears that the battle was far from won: just because something was called linguistic did not at all insure that it was indeed a good thing. A case in point is an approach to reading instruction that has taken regular education by storm and seems about to sweep special education as well. Its proponents (Goodman, 1976; Goodman & Goodman, 1979), who call reading "a psycholinguistic guessing game," suggest that because the main goal of reading is to derive meaning from print, we should teach children to go somehow directly from print to meaning, as skilled readers supposedly do. According to their position, the teacher should not correct a child who misreads dog as "cat." It is not such a bad error, they say--after all, since dogs and cats are both animals, the child has hit upon the correct category of meaning, and according to this instructional approach, it is general meaning, not the apprehension of any particular word, that should be rewarded. Moreover, they argue, attention to the phonology represented by the alphabetic characters would slow the reader down and make it harder for him to attend to meaning. In fact, a useful technique for teaching beginning readers, we are told by one practitioner of this approach (who apparently does not shrink from carrying it out in its most extreme form), would be to splash ink on the passage to be read and then to let the child practice reading by guessing what might have been hidden under the ink spots (Giordano, 1980).

The underlying assumptions of the psycholinguistic guessing game approach seem to be: first, that skilled readers do ignore the word and make little use of the phonology that is represented by the letters of the word, depending instead largely on guessing from the shape of the letters and the context to get at meanings; second, that readers can go faster that way; and third, that skilled readers have the kind of attentional control that permits them to determine by choice when to look at letters as representing the phonology and when to look at them only as visual shapes. All of these assumptions are questionable in our view, and, in any event, remain to be demonstrated. But perhaps the most misguided assumption of all, from my point of view, is that any reader should ever go directly from print to meaning.

ORTHOGRAPHIES: REPRESENTING UNITS OF LANGUAGE

I take it as given that in understanding language, whether written or spoken, one does not normally go directly to meaning. Rather, the listener or reader gets to the meaning via the language--that is to say, by dealing in distinctively linguistic ways with the units of the language (for example, phonological segments, words) and also the larger syntactic structures (sentences) they form. Surely, some kind of linguistic processing, however automatic, is necessary, for in language, as in everything else, there is no free lunch. Moreover, the processes that extract meaning from language are different in important ways from those that extract meaning from a picture. Perhaps one can go quite directly from a picture to one or another of its typically many meanings. I don't really know, and I suspect that no one else does either. But, whatever the processes by which we get meaning from a picture, the processes by which one gets it from language are different. Words and sentences are uniquely linguistic things, after all. A word is represented in a person's vocabulary as a string of abstract, meaningless phonological units, and its relation to meaning is arbitrary; there is absolutely nothing about a word that can possibly give its meaning "directly." As for a sentence, its meaning is even less directly available; surely, it is not to be had by summing the meanings of the constituent words. In some important sense, the meaning of a sentence is in its structure, and unearthing that meaning must depend on the use of uniquely grammatical devices--word inflections, word order, grammatical words (e.g., of, a); accordingly, the listener and the reader are both well advised to take account (we hope automatically and painlessly) of the appropriate grammatical structures and devices.

As ways of communicating messages, there is, then, an important difference between pictures and language (whether spoken or printed). Perhaps, as I suggested, there are pictures that do enable a viewer to "go directly to meaning." If that is an advantage, so be it. Indeed, I would add it then to another advantage that pictures have over language: they are often aesthetically more pleasing. But for the purpose of precisely conveying ideas, pictures are clearly inferior. How would you say, "The science of physics is far advanced," in pictures? But notice how easy it is to do that with language. Indeed, we can even do it with print, but only if the reader understands that the print represents the language.

All of the foregoing seems obvious enough, yet we are told by some that, because the main goal of reading is to derive meaning from print, which hardly needs saying, we should teach children to do that directly, which is a different matter altogether and badly wants contradicting. For if encouraging the child to go "directly to meaning" means anything at all, then it must be that we are being urged to teach the child that the print represents meanings, when, in fact, it represents the words of the language. And that does appear to be what we are being urged to do, when we are told--to take the example I used earlier--that the child who reads "cat" for dog is really on the right track. The basis for that misguided conclusion is that dog and cat are clearly related in some semantic way, so the fact that the child reads one when the other was written merely shows that his quick mind leaped immediately to the meaning and only missed it by a small amount. I would suggest, on the contrary, that this poor child has not the dimmest notion of what reading is about. The most likely explanation of his error is that he treated the word as if it were a picture, but being unable, of course, to determine precisely what it was a picture of, he looked at its general shape, remembered only that he had learned to associate that with some animal, and so, on being presented with dog, recalled another member of the set of animals he had seen represented. Such a child will never become an accomplished reader until he discovers--one hopes that a teacher might help him to discover--that the characters d o g are a phonological representation of a word. That word may have any or all of a variety of meanings to the reader. "That animal is a dog." "Why do you dog my footsteps?" "That movie is a real dog." But what stands fixed and firm is that the word is "dog" and that the print precisely represents it. (Imagine, by the way, how it might be that in reading a sentence, one would see a grammatical word like of. Would he go directly to its meaning? What is its meaning in isolation? Or, as I think plausible, would he read the word of and then hold it in some buffer until enough of the other words have accumulated to make it possible for him to apprehend the linguistic structure of which the word of is a part?)

But suppose all do agree that in reading a word the trick is to recover the word and then let the meanings follow as they normally do. There remains the question: how does (or should) the reader find the word? And here, too, we are often given advice that seems wrong-headed. I have in mind the frequently-made assertion that children should be taught to read words as wholes because that is what skilled readers are assumed to do. But, as I see it, the assumption that words should be read as wholes is either trivial or wrong, depending on just exactly what is meant. If reading a word as a whole means merely that one takes in a half dozen or so letters at a single fixation, then we are simply dealing with a well-known fact about optics, anatomy, and physiology, and not a prescription about how to read. Surely, all readers take in many letters (and most words) at a glance. But if, on the other hand, reading a word as a whole is meant to be a statement about how one reads, then it can only mean that the reader should not (does not) apprehend the internal phonological (or morphophonological) structure as represented by the letters, but rather should (or does) respond to some (always undefined) holistic characteristic. If that is what happens, however, then what kind of fix is the reader in when the word is itself not a whole--when, in fact, it has component parts? Take the words goodness and badness. If reading those words as wholes means anything at all, then it must mean that the reader does not apprehend the sublexical element--namely, "ness"--which is common to the two words, and that he therefore cannot appreciate that good is to goodness as bad is to badness. Or take walk, walks, and walked. To read those as holistically different from each other is to miss the critically important relations among them. It would seem, then, that to encourage a beginning reader not to take advantage of the phonological and morphophonological information in a printed word is to encourage him to miss a great deal of what is going on in the language and, inevitably, to become a poor reader.

Thus, my conception of the reading process begins with the seemingly obvious assumption that an orthography represents a language. It follows, then, that if we would understand what reading requires of a child, and especially why those requirements should so often be hard to meet, we must see exactly how the orthography represents the language, and why, given that kind of representation, it might be hard for the child to make the connection.

That is what has guided the research of my colleagues and me, and led us to pay particular attention to two critical aspects of the reading process. The first has to do with the reading (and understanding) of words: given a printed word, how does the reader (indeed, how should the reader) find in his lexicon the real word that the printed word represents? The second part has to do with the reading and understanding of sentences: given that the reader has got the words, how does he hold them until he can extract the meaning from the structures they form? In this paper, I will deal almost exclusively with the first: how one reads the words. I will be especially concerned to say why that might be difficult, and I will offer suggestions about how the teacher might make the task somewhat easier. I mean to take seriously the assertion that a writing system represents the language, for it is only when we understand this that we can see why certain kinds of difficulties might arise. So I will begin by describing various orthographies, including especially the one we use in English, with emphasis on the cognitive problems they present, especially to the beginning reader. Then, I will present evidence that these difficulties do, in fact, arise, and suggest how they have been misinterpreted. And, throughout, I will offer a few ideas about instruction that teachers will, I hope, find useful.

Picture writing, the earliest attempt to convey information for the eye, represented objects, events, and general meanings, rather than segments of language. By its very nature, however, it was open to different interpretations by different observers. A picture of archers meant by the artist to represent the hunt might, instead, have been interpreted by an observer as "archery," or "manliness," or "blood sport," or, indeed, as whatever other meaning the given observer might have associated with that picture. If we had not progressed beyond a pictographic system, therefore, we could communicate only vague, ill-defined areas of meaning.

Proper writing and reading may be said to have begun whenever it occurred to someone to convey a message, not by drawing a picture of some object or event, but by using optical patterns to represent the language. Though, as we will see, there are several ways to do that, the choices are really quite severely constrained. The first, and surely the most important, constraint has to do with a universal characteristic of language--to wit, that it is always made up of discrete units or segments (phones, phonemes, syllables, morphemes, words, phrases, sentences). The constraint on an orthography is that it must represent one or another set of those segments. (Imagine trying to read an orthography whose individual characters each represented a word and a half.) But there is a certain amount of choice as to just which segments will be represented. The most general aspect of this choice derives from a second universal characteristic of language: there are always two kinds of segments, meaningful (sentences, words, morphemes) and meaningless (phones, phonemes, syllables). Accordingly, some orthographies use their characters to represent meaningful segments, others one or another of the meaningless segments.

Let us, then, take a quick look at the several kinds of orthographies, trying in particular to see what various difficulties they might or might not present to the beginning reader. Among the meaningful units, we will here consider only the shortest unit, the morpheme, the unit most commonly represented. As for the meaningless units, we will consider the syllables, and also the constituent sounds, phones, and phonemes of which they are composed. The phonemes are, of course, the segments that are represented in the alphabetic orthography we use, but because there is so much confusion about what a phoneme is, and how it differs (or, indeed, whether it differs) from phonetic units and from the sounds of the language, I have included phonetic units and sounds as possible bases for an orthography.

The guiding principle of our search among the orthographies can be put very simply. Reading and writing are, by comparison with listening and speaking, relatively unnatural and derived. All speaker-hearers of a language are provided with a neurophysiology that normally functions naturally and automatically--that is, below the level of awareness, to cope with the structure of language (A. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). In contrast, the reader and writer must be something of a linguist--able, at the very least, quite deliberately to divide utterances into the constituent segments that are represented by the characters of the orthography. As we will see, the ease or difficulty with which that can be accomplished will depend, in large part, on the nature of the linguistic unit that the orthography represents.

ORTHOGRAPHIC REPRESENTATION OF WORDS

Among orthographies--true writing systems, that is, as distinguished from communication by means of pictures--are those that represent such meaningful units as morphemes or words. Certainly, the best known examples are Chinese and its adaptation in the Kanji part of Japanese. The exact ways in which the characters of these orthographies convey the Chinese and Japanese languages is complex (see, for example, Martin, 1972). For our purposes, however, it is sufficient, and sufficiently accurate, to say that the individual characters of the orthography, often referred to as logograms, represent morphemes (the shortest units of the language that have meaning) or words. Indeed, it does no real harm to the point I wish to make here to say that a character refers to a word. Of course, each logogram is decomposable into visually distinguishable parts (strokes), and these may be important in the recognition of the character, but they have no linguistic significance--they do not, for example, represent the sublexical phonological components of the word as the letters of our alphabet do. Logograms are used in English too--for example, the dollar sign or the arabic number 6--but they are the exception in our writing system.

From our point of view, the most important characteristic of a logographic writing system is that it presumably imposes a light cognitive burden on the beginning reader. To see why this is so, we again take account of the fact that any reader or writer must, at the least, be able to abstract from the utterances of a language exactly those units that the orthographic characters represent. (Like so many things that are important and seemingly obvious, this requirement is often unnoticed.) But if, as in the case of logographies, the unit is the word, then surely the cognitive task is relatively easy. Words are isolable units, after all, which is to say that they can be, and often are, produced outside the larger contexts (sentences) in which they typically occur. Nevertheless, studies have shown that very young children (Downing, 1971, 1972) are more than a little uncertain when asked (in effect) to abstract words from spoken sentences. But the difficulty is quite easily overcome (Engelmann, 1969). There remains, then, only the task of learning to associate a written character with the word it represents. That is simply paired-associate learning, and, up to a point, children are good at it.

There are, then, reasons for supposing that a logographic system should be quite easy for the beginning reader. Accordingly, we are not surprised to find evidence that perhaps it is. In a later section, I will outline that evidence. For the moment, let us simply ask: if a logographic orthography is relatively easy for children to master, why not teach them to read English as if each spelled word were a logogram? Why not, indeed, since we are often advised by educators (advocates of the "whole-word method," see Rosner, Abrams, Daniels, & Schiffman, 1981) to do precisely that (though not usually for the reasons given above). There are at least two reasons why not, and, precisely because we are so often urged to pretend that English should be read as if it were Chinese, I should take a moment to say what those reasons are.

The first reason why children should not be taught (or even permitted) to suppose that a spelled English word is a logogram is in the nature of the logographic system, and it is obvious: logographies are not as productive as the alphabet. That is, there is no way for a reader to read a morpheme whose associated logogram he had not previously seen and committed to memory. As a consequence, the reader of a logography must memorize thousands of characters, an assignment that will occupy him for many years. Even the Chinese have had to find ways out of this difficulty. Thus, for many of their characters--for most of them, if frequency of occurrence is taken into account--there are phonetic elements that lighten the memory load somewhat by providing indirect clues to pronunciation. In any case, a child who learns to read English words as if they were logograms will never be able to read a word he has never seen in print before. That much is surely obvious. Only slightly less obvious is the fact that, unlike the characters of the Chinese orthography, the letter strings formed by an alphabet are ill suited to be apprehended by overall shape or, indeed, by any means that does not take account of the distinct and distinctive letters. If we should be so misguided as to want children to read English words without appreciating their internal structure, we should, at the least, design an orthography that is more appropriate to that aim (Brooks, 1977).

The second reason has to do with differences between Chinese and Japanese, on the one hand, and English, on the other, differences that tend in the former cases, but not in the latter, to balance the inherent disadvantages of a logographic system with certain special advantages. Consider, in this connection, that there is in both Chinese and Japanese a great deal of homophony--many instances, that is, in which words that are phonologically the same are semantically different. Logograms nicely disambiguate these words and thus serve an important purpose. English does not have this characteristic to any considerable extent. We should also consider in this connection that Chinese has no inflections--for example, case or tense--so the user of a logographic system has only to associate logogram with word. There is no need to have a holistically different logogram for every inflected form of the word, nor is there, alternatively, any need to tax the reader-writer's linguistic ability by requiring him to mark the grammatical status of the word with some abstractly grammatical character that means, for example, "indirect object of the sentence." It is surely not trivial that in the Japanese adaptation of the Chinese orthography all grammatical inflections (and Japanese, unlike Chinese, does have these) are rendered phonologically in the Japanese syllabary (kana). English, of course, does have grammatical inflections which must be taken into account. Finally, there is in Chinese the special advantage that a logographic system can more easily be read across the several Chinese languages that are related but not mutually intelligible. We have no need for such an arrangement in English.

There are, then, two points to be made here. The first is that, yes, it is possible to represent a language orthographically with characters that refer not to the phonological constituents of words, but to the words themselves. But meanings are conveyed, in the orthography as in speech, by the words (including especially the grammatical words--*of*, *to*, *or*, etc.) and the larger grammatical structures they form. The second point is that, whatever special advantages a logography may have in Chinese or Japanese, it is ill suited to English. We have reason to be thankful that our English orthography is not logographic, and we should hesitate to design our reading instruction as if it were.

ORTHOGRAPHIC REPRESENTATION OF PHONOLOGIC UNITS

As we have seen, a logographic system is not, as it were, productive: readers cannot cope with a character-word correspondence they happen not to have seen before, but must rather learn a new character for every morpheme read. This is surely a great disadvantage, given that the number of morphemes in a language--hence the number of characters--runs into thousands. But when the characters of the orthography represent the meaningless units of the orthography that disadvantage is overcome: the phonological units are far less numerous than the words, and, once mastered, the system makes it possible for readers to cope with words they have not seen before, including even those newly invented words the language may have chosen to incorporate. Let us turn, then, to such orthographies, dividing them into two classes, according to the size of the phonological unit (the longer syllable or the shorter phone or phoneme) they represent.

Syllables and Syllabaries

Perhaps the best known example of a syllabary is the Japanese kana system. The linguistic unit is, strictly speaking, the mora, which is defined in temporal as well as ordinary syllabic terms, but we do not seriously misrepresent the matter if we regard it as a syllable and the orthography as a syllabary. In fact, there are two syllabaries for Japanese, the katakana, which is used for writing many imported foreign words, and the hiragana, for conveying grammatical inflections. There are 49 kana characters in each, corresponding to the same 49 syllables of the language.

What, then, is the cognitive burden that a syllabary imposes on a child? How difficult is it for him to abstract from his speech and from that of others the units that a syllabary represents? The answer to this question is to be found in part in the results of several studies (Calfee, Chapman, &

Venezky, 1972; Fox & Routh, 1976; Gleitman & Rozin, 1973; Liberman, 1971, 1973; Liberman, Shankweiler, Fischer, & Carter, 1974). These indicate that the young child comes more easily and more quickly to an explicit awareness of syllables than of the shorter phonological segments that an alphabet represents. The reasons for this are easy to see, once we understand how the processes of articulation and coarticulation merge the constituent phonetic segments into units of approximately syllabic length (A. Liberman et al., 1967). This is to say simply that, like words, syllables can be rather easily separated in the speech stream and pointed to, as it were, but most consonant constituents of a syllable cannot be made to stand alone (without an accompanying _schwa_). At all events, to the extent that a child must abstract from speech those units his orthography conveys, syllables present fewer difficulties than phones or phonemes.

But the research on how readily children become aware of syllable units only takes account of their ability to determine how many syllables there are in an utterance. It does not deal with their ability to find the exact boundaries. For a language like Japanese, in which syllables have a relatively fixed consonant-vowel structure ("Fuji," "Watanabe," "Mikimoto"), finding the boundaries poses no great problem. But where there is a great variety of syllable structures, as in English, the matter is considerably more difficult. Thus, even though we can easily perceive that a word like "federal" has three syllables, it is not that clear where the boundaries ought to be. We should also expect that a syllabary would be more troublesome as the number of different syllables increases, and then note in this connection that, in contrast to the small number of syllables in Japanese, there are thousands in English. The point, then, is that a syllabary might well have advantages for the reader, especially the child, but only in languages that have certain properties. English does not have those properties, and, in any case, it is not written with a syllabary.

Sounds, Phones, and Phonemes: Alphabets

We come now at last to the alphabetic orthography, the vehicle for the written form of English and, indeed, of most of the languages our students are likely to learn. The system has many advantages, especially for languages like English, but it also presents certain problems, both for the child who would learn to use it and for the teacher who would help him to do that. In reading an alphabet, as in reading a logography or a syllabary, the reader must be able quite explicitly to appreciate the relation between the orthographic character and the linguistic unit it represents. I have already made the point that this need not be very difficult for a logography or a syllabary. However, it can be quite difficult in the case of an alphabetic orthography (Liberman, 1971; Liberman et al., 1974). and it is so for reasons that we understand quite well. The essence of the problem can be put this way: though it is often said that an alphabetic orthography represents speech (or supposedly ought to in the ideal case), in fact, it is, and forever must be, an abstraction from speech. It does bear a regular relation to speech, barring a few egregious exceptions, but the nature of that relation is hard for the child to apprehend. To understand why, let us see in exactly what ways it is misleading to say that an alphabetic orthography represents the _sounds_ of speech.

Sounds. Alphabetic orthographies do not represent the sounds of speech. There are two senses in which this is so. One is obvious and quite trivial: the optical shapes of the letters do not portray the acoustic events, though they might well do just that if they were snippets of oscillograms or spectrograms. The other is not so well understood but far more important: the segmentation of the sound does not correspond directly to the segmentation indicated by the letters. Because of the way speech is normally articulated and coarticulated, information about several of the phonological and phonetic segments--the segments that are represented approximately by the letters of the alphabet--is transmitted simultaneously and on the same part of the sound. The consequence is that in a word like "big," for example, there is no acoustic segment corresponding to each letter segment. That is, it would be impossible to divide a recording of the spoken word "big" into three parts so that, when played back, one part would be "b," one part "i," and one part "g." In the syllable "big," there is but one piece of sound, and the three phonological segments that we write as b, i, and g have been more or less simultaneously encoded into it. This distinctively linguistic way of encoding the phonological segments into the sound is essential to the efficient perception of speech, for if each phonological segment were represented by a segment of sound, then communicating phonological structures at rates that range from 8 to 30 segments per second, as is normally done, would far overreach the temporal resolving power of the ear. As a result, the separate segments of the phonological message would merge in perception into an unanalyzable buzz. So, encoding several segments of the phonology into one segment of sound provides for an important gain in efficiency when one is listening to speech. But this gain exacts a price, for there is now a peculiar relationship between the phonological message and the acoustic signal that conveys it. Fortunately for the listener, however, he has access to a biologically specialized system that enables him effortlessly and automatically (though tacitly) to cope with the code and recover the message it conveys (for a fuller treatment of these matters, see A. Liberman, 1982; A. Liberman & Studdert-Kennedy, 1978; A. Liberman et al., 1967).

But the curious code that connects phonological structure to sound has two adverse consequences for the would-be reader. One is that it makes inordinately difficult the task of "reading" a spectrogram or, indeed, any other representation of the actual sounds of speech. Thus, it is not only true that alphabets do not, in fact, represents the sounds of speech, but, more important, it is just as well that they do not, for if they did, reading would be a slow and onerous business for us all.

The other consequence for the reader is that, for many of the segments of the language, there is no simple and direct way to demonstrate to him the relation between spelling and sound. If the teacher nevertheless undertakes to do this with a word like "big," she will be driven to isolate three sounds and in the process, she will unavoidably produce three syllables: "buh," "ih," and "guh." But they form a nonsense trisyllable, not the meaningful monosyllable that comprises the three phonological segments we spell as big.

None of this is to say that the phonological segments represented by the alphabet are fictions. Not at all. They are real enough and, as already indicated, are recovered at least tacitly by the listener as he processes the sounds of speech. But that processing is carried out by physiological mechanisms that appear tied to an acoustic input. If we would put speech into visible form and make it readable, we must, at the least, spell out the segmented form of the message by using the normal linguistic capacities of a human being to recover that form.

Phones. But suppose now that by paying careful attention to what we perceive when we listen to speech, we use the human being's linguistic ability to abstract from the acoustic signal the string of phonetic segments that it conveys, the phones. Now we are just one step removed from the sounds of speech. We have achieved a proper segmentation, and we can represent each perceived segment by an alphabetic character. Indeed, that is done in the phonetic alphabets that linguists use to transcribe as accurately as possible what they perceive when they listen to speech. But now we encounter another difficulty. It is that the wealth of phonetic information that the natural speech-perceiving mechanisms know how to use creates serious problems when, as in reading and writing, we short-circuit those natural mechanisms and put the information through the eye.

A phonetic transcription, that is, a transcription representing the phones of speech, preserves much surface information that is not represented in an alphabetic orthography. For example, a phonetically written orthography would reflect all the context-conditioned variations of speech both within words and across syllable and word boundaries. Thus, within words, the plural "s" after an unvoiced consonant, as in "cats," would be transcribed as *s*, but its counterpart after a voiced consonant, as in "dogs," would be transcribed as *z*, to reflect its pronunciation in that context. The stressed and unstressed forms of vowels would also be assigned different symbols instead of remaining the same as they do in *telegraph-telegraphy*. Similarly, the different pronunciations of the same consonant in different positions in a word, like the "t" in "tap" and in "pat," would demand different symbols because the careful listener could differentiate between them in the contexts of those two words.

The possibility that the recognition of such minute articulatory distinctions might actually detract from the broader requirements of efficient language representation becomes even more compelling when we see how context-conditioned variations of pronunciation across syllable and word boundaries would affect the phonetic transcription. For example, the final consonant in the word "bat" would be transcribed as *t*, but what we ordinarily consider to be the same consonant in the related word "batter" would have to be changed from *t* to *d*, in order to accurately reflect the manner change in our pronunciation of that segment from voiceless to voiced in the disyllabic context. Similarly, the contraction "what's" would be transcribed quite differently in the context of the sentence "What's he doing?" from its transcription in the context of "What's your choice?," where because of context-conditioned effects, it would be coarticulated with "your" to produce "Wuhchor choice?" in everyday spoken English and would therefore have to be transcribed that way in a phonetic rendition.

This brings us to another problem posed by a truly phonetic transcription, the question of what indeed is "everyday spoken English"? Idiolects, which would ordinarily be represented in a narrow phonetic transcription (e.g., a speaker's lisp, or difficulties with "l" and "r"), could perhaps be disregarded, but what about dialectical differences? Indeed, how would the received pronunciation be determined for purposes of devising an orthography? And would there need to be a different orthography, therefore, for English and American speakers of English?

It must now be apparent that it would be extremely difficult to apprehend a message that was conveyed by means of a narrow phonetic transcription. Though it has its uses for the phonetician whose very task it is to study these fine points of difference in speech, a phonetic transcription would usually give us as readers not only more information than we need, but actually, for our particular purposes, might often get in the way, by providing many data that we cannot efficiently use while hiding or obscuring other data that might have been helpful.

As it happens, although any literate adult can decode a transcription based on phones considerably more easily than he can decode a visual display of acoustic events, even highly trained phoneticians cannot read an unfamiliar text written phonetically with the same degree of fluency that they would show in reading the same passage written in our much maligned English orthography.

Thus far, in this necessarily brief discussion of options available for transcribing a language, we have touched upon the shortcomings, either in relation to cognitive load or to mismatch with our language, of a system using a meaningful unit, the morphemic unit of language, and also of several others, in which meaningless units, including syllables, sounds and phones, were the candidates for transcription. With these considerations in mind, we can now explore in somewhat greater detail the phoneme or morphophoneme, the meaningless segment that is used to represent the language in our alphabetic system.

Phonemes and Morphophonemes. Given that reading the sounds of speech is inordinately difficult and reading a proper phonetic transcription only slightly less so, what is it that an alphabet should represent if reading is to be as easy and fluent as possible? The relevant considerations are, I think, roughly as follows. We ask, first, how the words of the language are represented in your head and mine--in the lexicon every speaker has in his head. Certainly, they are not there as auditory templates, for, if they were, the speaker-listener would need a different lexicon for every different auditory shape that a word has as a consequence of variations in context, rate, linguistic stress, emphasis, idiolect, dialect, and goodness knows what else. Almost as certainly, words in our lexicons are not represented in narrow phonetic form, for in that case, too, we should have many lexicons, corresponding, again, to the numerous systematic variations that occur in response to many of the same factors that cause gross changes in auditory shape. Accordingly, it is altogether reasonable to suppose that some kind of systematic phonology, similar to what linguists like to talk about, does in fact exist as part of the normal person's language faculty. That is to say that your lexicon and mine are presumably organized in terms of phonological segments sufficiently abstract to stand above the many variations at the auditory and phonetic surfaces. Thus, you and I recognize that the word "telegraph" is the same word no matter what the idiolect or dialect (of English), and no matter what phonetic changes might have occurred because of a particular word that preceded or followed it in the sentence. Indeed, it is reasonable to suppose, at least in this case, that we tacitly command the rule that relates the phonetic structure of "telegraph" to the rather different phonetic structures of "telegraphy," and "telegraphic," and that the similar spellings are, accordingly, perfectly transparent.

When a person gets language by ear, then, the auditory and phonetic variations are processed automatically, yielding, finally, the more abstract form in which the word is contained in the listener's lexicon. Indeed, there is reason to believe that the more 'surfacy' variations in the auditory and phonetic domains actually provide important information, helping the system to isolate the words from the sentence contexts in which they appear and to identify them properly. But when we try to put language in by eye, then, as we have seen, difficulties arise if we begin with the (systematically) variable auditory and phonetic forms. To circumvent these difficulties, I should think we would want the words to be spelled in a way that precisely matches the quite abstract phonological structures in terms of which they are spelled in the reader's lexicon.

But there's the rub. For though we can be reasonably sure that the words in our lexicons are spelled quite abstractly, we don't really know exactly how abstractly. I suppose that, for most speakers of English, the phonetic "s" of "cats" and the phonetic "z" of "dogs" are represented the same in their lexicons, reflecting the underlying (morpho)phonological sameness of the plural, and I suppose the same is true for the phonetic changes that occur as a function of linguistic stress, as in the variations that are rung on a word like "telegraph." If those suppositions are correct, then it is, indeed, wise and proper that these words are spelled in the abstract form that immediately reveals to the reader what it is that they have in common. But what of the phonological alternations that make it sensible to keep the vowels the same in such pairs as heal-health, weal-wealth, and steal-stealth? One suspects that while some speakers of English comprehend those relationships, many others do not. Which brings us then to another difficulty we should have if we were trying to devise the ideal orthography: there are presumably great differences among speakers of the language in the way their lexicons are organized. To the extent that is so, the perfect orthography becomes impossible.

Given that every alphabetic orthography spells words quite abstractly, and given that this is as it should be, there remains a rather wide margin of choice as to just how abstract the system should be and precisely which abstractions it assumes the readers command (see Klima, 1972, and Venezky, 1970, for a more detailed discussion). For better or worse, English spelling is rather far out on the abstractness dimension, from which it follows that it must strain the linguistic sophistication of many who would read (and spell) it. The young child is especially likely to lack even the tacit knowledge that would rationalize so much of the spelling, and, as I mean to say in the next section, that creates a difficulty. But it is a difficulty that is not too hard to overcome, especially if the teacher truly understands its nature.

But perhaps the point to emphasize here is that no matter how abstract it may often be and how far or how close to a given reader's lexicon, the alphabetic orthography does, nonetheless, represent the internal phonological structure of the spoken word. Moreover, it does so by means of a remarkably economical set of only 26 symbols, which provide entry into the entire printed vocabulary of the language. To readers who understand and utilize the relationship between these symbols and the language, this orthography affords a unique advantage, certainly not available to the readers of a logography. Their advantage is that they can read words they have never seen before. They do not have to memorize the association between each symbol pattern and the word it represents before they can read it, as the logographic reader must.

LINGUISTIC SOPHISTICATION AND READING

In the light of the preceding discussion, we can turn again to the question of what children must know in order to learn to read. Beyond the obvious need to have some command of the language and the ability to discriminate the graphic symbols, the first requirement for beginning readers, in our view, is to acquire a certain amount of linguistic sophistication. The difficulty of acquiring the sophistication needed will, as I have said, vary with the language and the orthography. Having outlined the implications of the various orthographic options, we can now look more closely at the matter of linguistic sophistication and its role in reading English. For this purpose we would differentiate between two aspects of linguistic sophistication--phonological maturity and linguistic awareness (Liberman, Liberman, Mattingly, & Shankweiler, 1980).

Phonological Maturity

To the extent that English is written at the most abstract level, exemplified by the abstract linguistic relationships that rationalize the use of the same alphabetic characters for phonological segments that are phonetically quite different (as in cats and dogs, muscle-muscular, divine-divinity)--to that extent, it assumes an ideal reader who has assimilated the rules in terms of which that sort of spelling makes sense. That is, it assumes a reader who has, to some degree, what we have called phonological maturity.

Unfortunately, younger children may not have the degree of phonological maturity that an alphabetic orthography assumes. This is reasonably clear from the results of psycholinguistic research (Berko, 1958; Moskowitz, 1973) which suggests that young children are, indeed, quite immature phonologically and therefore not well-equipped to take full advantage of the more abstract aspects of the English orthography. Indeed, there is evidence from the invented spellings of preschoolers that young children actually do better as phoneticians than as phonologists (Read, 1975; Zifcak, 1977).

Luckily, while phonological maturity is of some importance in learning to read (and perhaps more so in learning to spell), it is not essential for the beginning reader. Our young phoneticians can learn to read, though perhaps a little awkwardly, mispronouncing a word here and there. We can help them along in these early stages of learning by controlling the vocabulary used in reading instruction (as is done in the so-called linguistic readers)--that is, by providing children with material that avoids the more difficult, less transparent alternations and only gradually increases the level of abstraction as the children show signs of understanding how the alphabet works. Indeed, it is probably experience in reading that, more than anything else, causes developing children to become sophisticated about the more abstract phonological regularities--for example, to realize how "magic" and "magician" are related. They do this by internalizing the phonological rules they induce from the orthographic transcription and by revising the representations of words in their lexicons accordingly. (Many, that is, will induce the rules; others may need to have the rules pointed out to them.)

Three points should be emphasized here. The first is that it is reasonable to suppose that the more one reads, the more one gains in phonological maturity. The second point is that this gain is possible only if, in reading, one attends to the relation between the printed word and the phonology of the spoken word, that is, if one reads analytically, not globally. One cannot develop this aspect of linguistic sophistication if one ignores the link between the orthography and the linguistic structures it conveys. And, finally, although it requires a linguistically sophisticated reader with a highly developed phonological sense to appreciate fully the extremely abstract way in which some of our words are written, entry into our orthographic system is quite possible without such a high level of that particular linguistic ability. More critical, in our view, for the beginner is the second aspect of linguistic sophistication, namely, the explicit understanding by the reader of the relation in segmentation between the orthography and speech (Liberman et al., 1974).

Linguistic Awareness

Until now we have been talking about the difference between a phonological representation and a phonetic one, and about the phonological maturity that allows the sophisticated reader to relate the two. Now we turn to another difference, that between the phonological domain in general (whether strictly phonetic or phonological) and the sound. In order to relate the phonological domain and the sound, the reader needs the second aspect of linguistic sophistication, what has been called "linguistic awareness" (Mattingly, 1972), that is, the explicit awareness of the segments that are represented by the orthography. As was noted earlier, it is clearly the case that the level of linguistic awareness required of a beginning reader will vary with the nature of the orthography, and, moreover, that entry into the alphabetic orthography, representing as it does the encoded sublexical units of speech, is more demanding than entry into, say, a logography, representing the more easily isolable word.

With all this in mind, we can consider once again the young child who is asked to read the word big. Let us propose that it is part of his speaking vocabulary, but that he has never before seen it in print. In our view, if the child is to map the three letters of the printed word onto the word he already knows (as he needs to do if he is to get from the print to the word), it will be of little use to him if all he is able to do is recognize the three letters, and, as he is often urged to do in "phonics" lessons, to "sound them out." In addition, he must also be helped to understand that the monosyllabic, seemingly indivisible word he knows has three segments, what those three segments are, and the order in which they occur. Unless he does know all that, given the impossibility of pronouncing the segments in isolation, he will produce something like "buh-ih-guh."

The point to be clarified here is that neither this child nor any other reader can recover speech from print on a letter-by-letter basis. What readers must do instead is to be able to put together the particular string of segments that, in ordinary speech, would be produced as a unit. The unit is commonly a syllable, but the number of letters that form a speakable unit can vary from one to as many as nine. In our view, learning to put together the letters into speakable units is a vital part of learning to read and one that may differentiate the fluent reader from the learner who is just beginning to see what an alphabetic orthography is all about (Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977).

Given these requirements of linguistic awareness, what can teachers do to ease the way for the beginner? As we see it, their first task is to help the child, as early as possible, to become aware of the segmentation of speech. Elsewhere (Liberman, Shankweiler, Blachman, Camp, & Werfelman, 1980), my colleagues and I have suggested several ways (pleasurable ways--they need not at all be the deadly drills that the "reading for meaning" advocates fear will turn children away from reading) in which this might be done, even in kindergarten, before the letters themselves are introduced. We have suggested beginning with nursery rhymes, word play, and word games, to be followed with any of the numerous activities specifically designed for this purpose by various educators such as Elkonin (1973), Engelmann (1969), and Rosner (1975). Actually, what may be most important at the start is simply to convince teachers that acquainting children with the segmental structure of speech is desirable--the teachers themselves will find countless and ingenious ways of doing it.

Once the children understand about segmental structure (first, perhaps, the words, then the syllables, and, finally, the phonemes), it becomes much easier to teach them how the alphabet transcribes the language. The teacher's next step would be to begin to teach the children the letters of the alphabet, their names, and sounds (see Slingerland, 1971, for an efficient and enjoyable way of doing this). As these are being taught and applied directly in reading and writing, the instruction need not, and, in fact, should not be limited to the traditional letter-by-letter phonics exercises (which are so often, and mistakenly, presented in disembodied lessons entirely separate from the reading class). They need not, that is, be limited to the practice commonly followed of urging the child to "Sound it out; say it faster; blend it." Such a practice may be defensible in the early stages of reading instruction, but only when used with letters like s, m, and n, which can be sounded without the accompanying schwa. It is quite unsuitable, however, for the highly encoded stop consonants (b,d,g,p,t,k) where speed of production will do little to promote blending and continued failure to blend the unblendable may, indeed, turn the child away from reading. We have advocated, instead, various ways in which the teacher can make use of consonant-vowel and vowel-consonant combinations in order to lead the child to map the letters to the phonology and learn, thereby, how to really read words (Liberman et al., 1980). (I hasten to add that these methods are not new--many thoughtful teachers have probably been using similar procedures since reading began. Our aim is simply to encourage their wider use by providing a reasonable motivation for doing so.)

MEANING AND THE WORD IN BEGINNING AND SKILLED READING

The basic task of the readers of any orthography is to get from the printed word to the appropriate word in the lexicon. Though I would, of course, agree that the apprehension of meaning is the ultimate aim of reading, I would wish to emphasize what seems an obvious (but often neglected) fact--that readers cannot apprehend the intended meaning of a sentence unless and until they have apprehended its constituent words. The last question we will address is how this requirement might affect beginning and skilled readers.

The Beginning Reader

I have gone to considerable lengths to show that because the particular speech segment represented by the alphabetic orthography is sublexical and difficult to isolate, the cognitive demands on beginners will be greater (and the task of the teacher harder) and that English further compounds the difficulty for them by the highly abstract way in which it often represents the language. In consequence, I have proposed, as others have (Gleitman & Rozin, 1977; Rozin & Gleitman, 1977), that learning to read will be harder for beginning readers of English than for beginners of Chinese, where the segment to be extracted from the speech stream is the easily isolable word, where any subsequent analysis of the phonological structure of the word is minimal, and where simple paired-associate memory of symbol and word is sufficient for mastery.

Many educators currently concerned with reading apparently disagree. To cite a recent example (Rosner et al., 1981), some would have us believe, instead, that reading is basically "a process of association" and that the problem of the poor beginning reader of English is "symbolization and association." In that view, the dyslexic "experiences difficulty in the association of common experiences and the symbols representing them." Their recommendation for reading instruction is that it "should be meaning-based with a modified language experience approach using content-materials as a vehicle. Word learning in the experience approach should be a whole word procedure for pedagogy."

Since similar views are so widely held, it might be useful to consider them here in some detail. First, is reading a process of association? Well, of course, it can be (though it would be the association of symbols with words, not with experiences--to my knowledge, no orthography uses its symbols to represent experiences). That is, there is nothing to stop a learner from approaching an alphabetic orthography as if it were a logography. Beginning readers of English can, if they choose or are taught to do so, approach their task just as Chinese children do. That is, they can treat the alphabetically written word as if it were a logogram--a graphic pattern like the dollar sign, which bears no relation to the internal segmental structure of the word "dollar." In other words, they can, indeed, adopt a "whole-word" strategy--learning to read by associating each pattern of letters with the word it represents, and presumably using the context to guess at the identity of graphic patterns they have not yet memorized. But by so doing, they will, of course, lose all the remarkable benefits of the alphabetic system. Like Chinese children learning logograms, they will begin to amass a collection of memorized graphic patterns and their associated words. They will not be able to use the alphabet in the way it was intended, to help them to apprehend new words. For them, a new word will simply be a new graphic pattern to be paired with an associated word, memorized, and added to an ever increasing collection of memorized symbol-word associations. As the collection gets larger, what small advantage there was in starting out this way should certainly soon begin to be lost.

It must be added, in good conscience, that, despite being taught by a whole-word method, some children sooner or later do discover the alphabetic principle on their own; that is, they themselves notice the relationship between how the word is spelled and its phonological structure, and begin to use that knowledge to good effect. We take this as the triumph of their native linguistic ability over the efforts of the whole-word method to keep the principle hidden from them. But what about the many children in our schools who are poor readers or even nonreaders? Is their problem really a defect in associative ability? Since our schools have been introducing reading by a kind of whole-word method for many years (by teaching children to memorize an introductory set of symbol-word associations to be triggered by picture- and story-context), one must wonder whether the problem of many of our poor readers was that they continued doggedly with the whole-word, logographic strategy, never managing to see the alphabetic principle on their own, and thus falling farther and farther behind their more perceptive classmates or finally giving up.

In any event, I would seriously question whether the poor reader's problem is one of symbolization or association. I know of no evidence that would suggest that this is really the case, and considerable evidence to the contrary. For example, learning disabled children who have never been able to master an alphabetic orthography readily learn to pair Chinese-like characters with their associated words and then to read off strings of them that have been arranged to form sentences (Rozin, Poritsky, & Sotsky, 1971). Moreover, a recent study (House, Hanley, & Magid, 1980) has shown that even retardates with a mental age of five or even less, who had never been able to learn to read, can be taught to identify and remember 200 or more pseudologograms and then to read them correctly when they appear in sentence form. They are simply taught to pair a visual pattern with a word and to memorize the association between the two. Surveys of dyslexia research also abound with many studies which strongly demonstrate that disabled readers have no difficulty at all in paired-associate memory (see Vellutino, 1979, for a recent review). In contrast, poor analytic linguistic abilities (as in phoneme and even syllable segmentation) are consistently found to be related to and predictive of poor reading achievement (Blachman, 1981; Calfee, Lindamood, & Lindamood, 1973; Goldstein, 1976; Golinkoff, 1976; Liberman & Mann, 1981; Lundberg, Olofsson, & Wall, 1980; Treiman & Baron, 1981).

Now what about the notion that "reading instruction should be meaning-based with a modified language experience approach"? As I have said earlier, it seems obvious that the meaning of a word cannot be apprehended without first apprehending the word itself and that the meaning of sentences and paragraphs cannot be apprehended without first apprehending their constituent words and grammatical structures.

Here it is useful to emphasize again that a word is something apart from its meanings. One does not have to know the meaning of a word in order to be able to read it (or to say it, for that matter). One can read a word like *blastoderm* but not know its meaning and therefore have to look it up in a dictionary or ask someone for its meaning. On the other hand, one can read a word like *club* and know several meanings for it, but have to determine from the context which meaning the author intended. In the first case, one must depend on a dictionary or a knowledgeable person for the meaning; in the second case, one can use one's own knowledge to arrive at the meaning. But in either case, before one can get to the meaning of the word represented by the print, one must first get from the print to the word. And modified or not, a language experience approach will not inform our readers how to get from the print to all the new words they encounter.

The Skilled Reader

So much for the beginning reader. What of the skilled reader? The received view in educational circles appears to be that once you are a skilled reader, you have found some miraculous way of discovering what the writer said, without first recovering what he actually said, and that the less you get of the information provided you by the print, the more skilled you are, because you are faster (Goodman & Goodman, 1979; Smith, 1973). As for the psychological literature on reading, much of the discussion there swirls around whether you arrive at the information in the print by an acoustic code, by a phonetic code, by a visual code, or by some interactive method in which you rely heavily on context but do examine words as you need to do so.

I would say again in response to all this that the acoustic signal is not represented by the alphabetic orthography, so all talk of an acoustic code is irrelevant. As to a phonetic code, the exact phonetic information, as we have seen, is also not represented in the alphabetic orthography and, indeed, there are few instructions in the print as to exactly how to produce it. It is just as hard to see how a visual code would work. The linguistically relevant information is not given by the overall optical configuration of the word nor by the optical shapes of the letters (the ascending, descending, diagonal, or circular characteristics of the squiggles on the page). As to the interactive approach, its proponents seem to be suggesting that in reading a passage, the skilled reader can go along deciding whether to read a word or whether to use the context to guess at it. In my view, if you are a skilled reader, your reading of words is automatized; you cannot keep yourself from reading the words. You cannot go along deciding whether you will read the word or will instead guess its identity from the context. You do use the context on occasion, of course--for example, when you are jarred by a conflict between a word you have read and the meaning of the rest of the words in the sentence or perhaps to determine the meaning of a word you have read. But in both cases, you will have read the words. This is not to say that a skilled reader cannot skim through a book or passage, reading a word here and there, or that he cannot skip over the long polysyllabic, hard-to-pronounce names in Russian novels. But in neither case is he using the context to get at the word. In the first case, he is actually reading words to get the meaning and in the second case, he is simply not reading.

Now to get back to what the skilled reader does do when he reads. Since an alphabetic orthography represents linguistically relevant aspects of the internal structure of the word, the reader, no matter how skilled he is, misses a lot if he ignores it.

What is he missing? The internal structure of the word can provide information about its derivational status and the constituent morphemic elements of polymorphemic words. It can provide information also about its grammatical status--for example, the tense, case, and number of the words and the effect of prefixes and suffixes on them. If you are going to get all that information from the printed word, you are well advised, in reading the word, to apprehend the internal structure which is, in fact, represented by the letters. Even if you have seen the word a million times, you nonetheless need to take account of its structure, if you are properly to understand what you read.

In this section, I have tried to answer three questions about what the skilled reader does. First, does he go directly to meaning or does he read the words? Second, if he bothers with the words at all, does he guess at what they might be from the context and pay attention to them only when all else fails? And third, does he read words as wholes or does he pay attention to their internal structure? In my view, it is the poor reader (and the beginner), not the skilled one, who attempts to go directly to meaning, who guesses frequently at words from the context, and who reads words as wholes. The skilled reader, in contrast, attends to the words and their phonological structure, and guesses only rarely (see Gough & Hillinger, 1980; Perfetti, Goldman, & Hogaboam, 1979).

In sum, if they are to make best use of an alphabetic orthography, both the skilled reader and the beginner must apprehend the internal structure of the word. The skilled reader does it quite automatically, and beginners, though it may be difficult for them, should be given directed instruction toward that end from the start. That is, they should be instructed from the start as to just how the orthography represents words. They should not be taught as if reading were a matter of associating a visual shape with a meaning or as if reading can be mastered without learning how to use an alphabetic orthography properly, or as if it should depend heavily on guessing from shape and context. As I have tried to show, such notions surely go against all we know about language, the orthography, and the reading process.

REFERENCES

Berko, J. The child's learning of English morphology. _Word_, 1958, _14_, 150-177.

Blachman, B. A. _The relationship of selected language measures and the McCarthy Scales to kindergarten and first-grade reading achievement._ Unpublished doctoral dissertation, University of Connecticut, 1981.

Brooks, L. Visual pattern in fluent word identification. In A. S. Reber & D. L. Scarborough (Eds.), _Toward a psychology of reading: The proceedings of the CUNY conference_. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1977.

Calfee, R., Chapman, R., & Venezky, R. How a child needs to think to learn to read. In L. W. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, 1972.

Calfee, R. C., Lindamood, P., & Lindamood, C. Acoustic-phonetic skills and reading--Kindergarten through twelfth grade. *Journal of Educational Psychology*, 1973, *64*, 293-298.

Downing, J. Children's developing concepts of spoken and written language. *Journal of Reading Behavior*, 1971-72, *4/1*, 1-19.

Elkonin, D. B. U.S.S.R. In J. Downing (Ed.), *Comparative reading*. New York: Macmillan, 1973.

Engelmann, S. *Preventing failure in the primary grades*. Chicago, IL: Science Research Associates, Inc., 1969.

Fox, B., & Routh, D. K. Phonemic analysis and synthesis as word attack skills. *Journal of Educational Psychology*, 1976, *68*, 70-74.

Giordano, G. Three criteria for judging initial reading instruction. *Academic Therapy Quarterly*, 1980, *15*, 291-297.

Gleitman, L. R., & Rozin, P. Teaching reading by use of a syllabary. *Reading Research Quarterly*, 1973, *8*, 447-483.

Gleitman, L. R., & Rozin, P. The structure and acquisition of reading: Relations between orthographies and the structure of language. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading: The proceedings of the CUNY Conference*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

Goldstein, D. M. Cognitive-linguistic functioning and learning to read in pre-schoolers. *Journal of Educational Psychology*, 1976, *68*, 680-688.

Golinkoff, R. Phonemic awareness skills and reading achievement. In F. B. Murray & J. J. Pikulski (Eds.), *The acquisition of reading: Cognitive, linguistic, and perceptual perspectives*. Baltimore: University Park Press, 1976.

Goodman, K. S. Reading: A psycholinguistic guessing game. In H. Singer & R. B. Ruddel (Eds.), *Theoretical models and processes of reading*. Newark, Del.: International Reading Association, 1976.

Goodman, K. S., & Goodman, Y. Learning to read is natural. In L. Resnick & P. Weaver (Eds.), *Theory and practice of early reading*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1979.

Gough, P. B., & Hillinger, M. L. Learning to read: An unnatural act. *Bulletin of the Orton Society*, 1980, *30*, 179-196.

Hammill, D. D. Training visual perceptual processes. *Journal of Learning Disabilities*, 1972, *5*, 552-559.

Helfgott, J. Phoneme segmentation and blending skills of kindergarten children: Implications for reading acquisition. *Contemporary Educational Psychology*, 1976, *1*, 157-169.

House, B. J., Hanley, M. J., & Magid, D. F. Logographic reading by TMR adults. *American Journal of Mental Deficiency*, 1980, *85*, 161-170.

Klima, E. How alphabets might reflect language. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, MA: MIT Press, 1972.

Liberman, A. On finding that speech is special. *American Psychologist*, 1982, *37*, 148-167.

Liberman, A. M., Cooper, F. S., Shankweiler, D., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, *74*, 431-461.

Liberman, A., & Studdert-Kennedy, M. Phonetic perception. In R. Held, H. Leibowitz, & H-L. Teuber (Eds.), *Handbook of sensory physiology, Vol. VIII, Perception*. Heidelberg: Springer-Verlag, 1978.

Liberman, I. Y. Basic research in speech and lateralization of language: Some implications for reading disability. *Bulletin of the Orton Society*, 1971, *21*, 71-87.

Liberman, I. Y. Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society*, 1973, *23*, 65-77.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. Orthography and the beginning reader. In J. Kavanagh & R. Venezky (Eds.), *Orthography, reading, and dyslexia*. Baltimore: University Park Press, 1980.

Liberman, I. Y., & Mann, V. Should reading instruction and remediation vary with the sex of the child? In A. Ansara, M. Albert, A. Gallaburda, & N. Gartell (Eds.), *The significance of sex differences in dyslexia*. Towson, MD.: Orton Society, 1981.

Liberman, I. Y., Shankweiler, D., Blachman, B. A., Camp, L., & Werfelman, M. Steps toward literacy. In P. Levinson & C. Harris Sloan (Eds.), *Auditory processing and language: Clinical and research perspectives*. New York: Grune & Stratton, 1980.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, 1974, *18*, 201-212.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. Scarborough (Eds.), *Toward a psychology of reading: The proceedings of the CUNY conference*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

Lundberg, I., Olofsson, A., & Wall, S. Reading and spelling skills in the first school years, predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology*, 1980, *21*, 159-173.

Martin, S. Nonalphabetic writing systems: Some observations. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, MA: MIT Press, 1972.

Mattingly, I. G. Reading, the linguistic process, and linguistic awareness. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationships between speech and reading*. Cambridge, MA: MIT Press, 1972.

Moskowitz, B. A. On the status of the vowel shift in English. In T. Moore (Ed.), *Cognitive development and acquisition of language*. New York: Academic Press, 1973.

Perfetti, C. A., Goldman, S. R., & Hogaboam, T. W. Reading skill and the identification of words in discourse context. *Memory & Cognition*, 1979, *7*, 273-282.

Read, C. *Children's categorizations of speech sounds in English*. NCTE Research 17, ERIC, 1975.

Rosner, J. *Helping children overcome learning difficulties*. New York: Walker and Co., 1975.

Rosner, J., & Simon, D. P. The Auditory Analysis Test: An initial report. *Journal of Learning Disabilities*, 1971, *4*, 384-392.

Rosner, S. L., Abrams, J. C., Daniels, P. R., & Schiffman, G. B. Dealing with the reading needs of the learning disabled child. *Journal of Learning Disabilities*, 1981, *14*, 436-448.

Rozin, P., & Gleitman, L. R. The structure and acquisition of reading II: The reading process and the acquisition of the alphabetic principle. In

A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading: The proceedings of the CUNY conference*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

Rozin, P., Poritsky, S., & Sotsky, R. American children with reading problems can easily learn to read English represented by Chinese characters. *Science*, 1971, *71*, 1264-1267.

Slingerland, B. H. *A multi-sensory approach to language arts for specific language disability children: A guide for primary teachers*. Cambridge, MA: Educators Publishing Services, Inc., 1971.

Smith, F. *Psycholinguistics and reading*. New York: Holt, Rinehart, and Winston, 1973.

Treiman, R., & Baron, J. Segmental analysis ability: Development and relation to reading ability. In G. E. MacKinnon & T. G. Waller (Eds.), *Reading research: Advances in theory and practice* (Vol. 3). New York: Academic Press, 1981.

Vellutino, F. R. *Dyslexia: Theory and research*. Cambridge, MA: MIT Press, 1979.

Venezky, R. L. The structure of English orthography. The Hague: Mouton, 1970.

Zifcak, M. *Phonological awareness and reading acquisition in first grade children*. Unpublished doctoral dissertation, University of Connecticut, 1977.

PHONETIC FACTORS IN LETTER DETECTION: A REEVALUATION*

Adam Drewnowski+ and Alice F. Healy++

Abstract. Three experiments in which subjects searched for the letter e in printed text were conducted to examine the effects of phonetic factors in silent reading. In Experiment 1, subjects made more errors on silent es than on voiced es, but silent es always occurred at the ends of words, whereas voiced es occurred in the middle of words. In Experiment 2, all instances of the letter e occurred in the penultimate location in the words, and no effects of letter voicing were obtained. In Experiment 3, subjects made more errors on es in unstressed syllables than on es in stressed syllables in three-syllable words. However, this effect occurred only for es in the second and third syllables and only for the more common words. All three experiments yielded large effects of word frequency, which were reduced in passages printed in alternating typecase. It was concluded that letter detection is affected by syllable stress but not by letter voicing and that the stress effect depends on whether the subject is able to form reading units at the syllable level.

There is much evidence that phonetic recoding of text occurs in the course of silent reading. One of the most influential studies (Corcoran, 1966) demonstrated that subjects searching for instances of the letter e in printed text made more errors on words in which e was silent (as in the word time) than on words in which it was pronounced (as in the word well). The

---

*Also in Memory & Cognition, 1982, 10, 145-154.

+The Rockefeller University.
++University of Colorado.

Acknowledgment. This research was supported in part by a grant from the Spencer Foundation to Yale University, National Science Foundation Grants BNS80-00263 to Yale University and BNS80-25020 to the University of Colorado, and National Institute of Child Health and Human Development Grant HD01994 to Haskins Laboratories. Adam Drewnowski was a Postdoctoral Fellow at the University of Toronto under Grant APA 146 from the National Research Council of Canada during the conduct of these experiments. Alice F. Healy was supported by a Senior Faculty Fellowship from Yale University during the initial preparation of this manuscript. We are indebted to Doris Glavanov, Mary LaRue, Michele Noe, and Cheryl Wright for help with the conduct and analysis of the three experiments and to W. K. Estes, B. B. Murdock, and G. Wolford for helpful discussions about this research. Parts of this paper were presented at the Sixth Annual Interdisciplinary Conference, Park City, Utah, January 19-23, 1981.

common interpretation of this result, also observed by other investigators (e.g., Chen, 1976; Coltheart, Hull, & Slater, 1975; Locke, 1978; Mohan, 1978), is that subjects silently reading paragraphs of text scan the acoustic image of a word along with the visual stimulus. However, in normal English prose of the type used by Corcoran, the voicing of the letter *e* is typically confounded with a number of other factors. For example, silent *e*s are often found in terminal or penultimate locations within words (e.g., *some*, *states*), and many occur in frequent function words (e.g., *have*) or in morpheme suffixes (e.g., *asked*). Each of these variables has been shown to influence the number of errors in letter-detection tasks: More errors have been found when the target letter occurred at the end of words (Corcoran, 1966; Smith & Groat, 1979), in frequent words (Healy, 1976, 1980), in function words (Drewnowski & Healy, 1977; Schindler, 1978), and in some morpheme suffixes (Drewnowski & Healy, 1980). In the present study, we used specially prepared texts that control for these variables in order to determine whether voicing of the target letter has a residual effect on the detection task. Our study was intended as a systematic reexamination of Corcoran's (1966) silent-*e* effect in an attempt to specify the nature and to determine the boundary conditions of the phenomenon.

Our previous research with the letter-detection task (Drewnowski & Healy, 1977, 1980; Healy, 1976, 1980) has shown that subjects miss letters most often in the most common words, suggesting that frequent words may often be perceived in terms of units that include more than one letter. According to our frequency-dependent unitization model (see, e.g., Drewnowski & Healy, 1977), the constituent letters of the most frequent English words tend, in effect, to be concealed within the word, since they never reach the level of identification in the course of fluent reading.

In our view (Drewnowski & Healy, 1977), reading involves processing in parallel of units at various levels of the linguistic hierarchy: letters, letter groups, words, or phrases. The ease of unit formation depends on the frequency and spatial predictability of letter sequences (Drewnowski & Healy, 1980), whole word frequency (Healy, 1976, 1980), and the syntactic constraints of text (Drewnowski & Healy, 1977). We have assumed that once processing at some higher level is complete, subjects move to the next location in the text without necessarily completing the processing at the letter level, at least not to the point of letter identification. Such incomplete processing at the letter level does not interfere with the comprehension of text, but it may account for the missing-letter effect, which we have observed for the most common suffix morphemes (Drewnowski & Healy, 1980) and for the most frequent words (Drewnowski & Healy, 1977; Healy, 1976, 1980).

This model leads us to predict that the type of phonetic effects observed will depend on whole-word frequency. If the more frequent words are indeed processed in terms of syllable or word units, rather than letter units, then phonetic effects at the letter level should be relatively unimportant. For common words, phonetic effects at the letter level, as exemplified by the difference in error rates between silent and pronounced *e*s, may be less important than phonetic effects at the syllable level, as exemplified by a difference in error rates between *e*s in stressed and unstressed syllables. Thus, the phonetic effects involving syllable stress may be more closely aligned to the postlexical phonological codes investigated by Foss and Blank

(1980) than to their prelexical phonetic codes. However, we envision phonetic units at the syllable level, as well as at the word level. This notion of a hierarchy of phonetic units, analogous to the hierarchy of visual units, is similar to that proposed earlier by LaBerge and Samuels (1974).

In the present study, we examined phonetic effects at the syllable level as well as at the letter level as a function of whole-word frequency. Specifically, we used both common and rare words to investigate the subjects' ability to detect the letter e in syllables that did or did not carry the primary word stress. In addition, we manipulated visual and linguistic features of text to determine the extent of their interactions with word frequency and phonetic factors in the course of silent reading. Understanding these multiple interactions should help us extend our theoretical conception of the reading process.

EXPERIMENT 1

The first experiment was designed to re-examine the acoustic scanning hypothesis (Corcoran, 1966) and our unitization model. The voicing of the target letter e (silent vs. pronounced) and the linguistic class of the target word (function vs. content) were independently varied. The voicing of the letter e deliberately covaried with its location within the word: Silent es were always terminal, whereas pronounced es always occurred in the interior of test words, as is typically the case in English. Also, because English function words are normally more frequent than content words, the function words that were selected as test words were, on the average, more frequent than the test content words.[1]

To determine the contribution of perceptual features and of the syntactic/semantic context to performance on the letter-detection task, the subjects were tested on four different passages. In addition to a standard prose passage, the subjects were presented with a nonsense passage of scrambled words, and with a mixed-case prose passage in which alternating letters were typed in uppercase. Although such manipulations should not affect the acoustic scanning of the search text, they are expected to impede the formation of reading units larger than the letter (mixed-case passage) or reading units larger than the word (scrambled-word passage) and consequently should influence the incidence of letter-detection errors. A fourth passage of meaningless and unpronounceable letter strings containing instances of the letter e in the same locations as the corresponding words in the prose passage was included to determine the effects of target location on task performance. (See Drewnowski & Healy, 1977, and Healy, 1976, 1980, for similar passage manipulations.)

Method

Subjects. Eighty-two students at the University of Toronto served as volunteer subjects in a group experiment, which was conducted in the classroom.

Design and materials. Four 100-word passages, typed on separate sheets of paper, were constructed for the present experiment. The first passage, hereafter referred to as the "prose standard-case" passage, contained 16 test function words (see Schindler, 1978, for definition) and 16 test content words, all of which contained exactly one instance of the letter *e*, along with 68 filler words, none of which contained the letter *e*. All test words were either one or two syllables long and varied from three to seven letters in length. Eight of the function words were judged to possess a pronounced *e*, which occurred in some intermediate position of the word: *they*, *their*, *them*, *her*, *after*, *under*, *over*, *himself*. The mean frequency of usage of these words (from Kučera & Francis, 1967) was 1,841 per million words of text. The other eight function words were judged to possess a silent *e*, which always occurred at the end of the word: *are*, *have*, *those*, *one*, *above*, *like*, *since*, *whose*. The mean frequency of these words was 1,868. The 16 content words were similarly divided into eight with pronounced *e*s (*well*, *men*, *years*, *get*, *very*, *later*, *given*, *power*), with a mean frequency of 659, and eight with silent *e*s (*time*, *use*, *make*, *home*, *office*, *little*, *middle*, *course*), with a mean frequency of 650. Mean word frequency across the voicing conditions was approximately equal (pronounced: 1,250; silent: 1,259).

The second passage, hereafter referred to as the "prose mixed-case" passage, was identical to the prose standard-case passage, except that alternating letters were typed in upper- and lowercase. There were two versions of this passage. In one version ("even"), even letters were capitalized, whereas in the other version ("odd"), odd letters were capitalized. Half the subjects were shown the even version of the prose mixed-case passage and half were shown the odd version, so that the incidence of lowercase and uppercase *e*s would be equated across test words and across subjects.

The third passage, hereafter referred to as the "scrambled-word" passage, was derived from the prose passage. The order of the 32 test words embedded within the paragraph of text was the same as in the prose passage, but the order of the remaining 68 filler words (none of which contained the letter *e*) was now randomized so that the passage no longer made sense. Whenever two test words occurred together in the prose passage (e.g., *little time*), they were separated in the scrambled-word passage by a filler word (e.g., *little who time*), but otherwise, the test words retained their original positions. Such manipulations were intended to minimize the presence of syntactically correct units in an otherwise meaningless passage.

The fourth passage, hereafter referred to as the "scrambled-letter" passage, was also derived from the prose standard-case passage. The letters in each of the 20 consecutive 5-word strings in the prose standard-case passage were now randomized to produce meaningless letter strings that corresponded both in length and in the location of the letter *e* within the string to the words of the prose standard-case passage. The location of the "words" on the page, the paragraph format, and the punctuation marks were the same as in the prose standard-case passage. The first lines of the four passages are shown in Table 1.

Table 1

First Lines of the Four Search Passages Used in Experiment 1

Prose Standard Case:

Men who work very long hours pass too little time at home.

Prose Mixed Case (Even):

mEn WhO wOrK vErY lOnG hOuRs PaSs ToO lItTlE tImE aT hOmE.

Scrambled Word:

Men his with very only cloud pass little who time an home.

Scrambled Letter:

Mer wlo vrny weok hnog loirs posi tao hutlte tsme ah twse.

Each passage was typed on a separate sheet of paper. The four passages, arranged in all 24 possible sequences and preceded by a page of instructions to subjects, were stapled together into a booklet. The booklets were distributed according to a fixed rotation so that passage order was approximately counterbalanced across subjects.

Procedure. The subjects were instructed to read each passage silently at their normal reading speed and to circle each instance of the target e. The subjects were told that if they ever realized that they had missed a target, they should not retrace their steps to encircle it. They were also told that they were not expected to detect all the es, so they should not slow down their reading speed in order to be overcautious about encircling the es. The subjects were told to read the passages in the order in which they were stapled together, and to go on to the next passage as soon as they had finished the preceding one.

Results

The results are summarized in Table 2, which includes for each of the four passages the mean error percentages (and standard errors of the mean) as a function of the voicing of the target letter and the class of the test word.

Table 2

Means (and Standard Errors) for Error Percentages as a Function of Passage Type, Voicing of the Target Letter, and Word Class in Experiment 1

	Word Class			
	Function		Content	
Passage Type	Pronounced	Silent	Pronounced	Silent
Prose Standard Case	12.63 (1.87)	31.38 (2.87)	9.88 (1.62)	16.88 (1.75)
Prose Mixed Case	11.38 (2.12)	23.88 (2.75)	9.63 (2.25)	16.88 (2.12)
Scrambled Word	12.75 (1.87)	27.88 (2.62)	8.63 (1.25)	16.00 (2.00)
Scrambled Letter	7.25 (1.50)	16.50 (1.87)	4.25 (1.25)	9.88 (1.62)

More errors occurred on silent than on pronounced *es*, $F(1,81) = 92.0$, $p < .01$, and more errors were made on function words than on content words, $F(1,81) = 74.4$, $p < .01$. In addition, the difference in error rates between the pronounced and the silent *es* was greater for function words than for content words; there was a significant interaction between word class and voicing, $F(1,81) = 29.4$, $p < .01$.

The subjects performed similarly on both the prose standard-case and scrambled-word passages (mean overall error percentages: 17.7 for prose standard-case; 16.3 for scrambled-word) and were somewhat more accurate on the prose mixed-case (15.4) and considerably more accurate on the scrambled-letter (9.5) passages, $F(3,243) = 9.3$, $p < .01$. The difference in error percentages between function words and content words was greater for the prose standard-case and the scrambled-word passages than for the prose mixed-case and the scrambled-letter passages. The interaction between word class and passage type, $F(3,243) = 3.1$, $p < .05$, supports our view that intact word units are necessary for the missing letter effect. The difference in error percentages between silent *es* and pronounced *es* also depended on passage type; the interaction between voicing and passage type was significant, $F(3,243) = 2.8$, $p < .05$. Nevertheless, even in the nonsense scrambled-letter passage, the difference between "pronounced" and "silent" *es* was significant, at the equivalents of both function word, $t(81) = 3.7$, $p < .01$, and content word, $t(81) = 2.3$, $p < .01$, locations. Since "pronounced" *es* always occurred in the middle and "silent" *es* always at the end of the nonword letter strings, these findings suggest that error rates in the letter-detection task may be strongly influenced by target location.

Discussion

The present results are consistent with Corcoran's (1966) finding that subjects searching for instances of the target letter *e* made more detection errors on silent than on pronounced *es*. However, the results are equally consistent with our previous reports (Drewnowski & Healy, 1977, 1980; Healy, 1976, 1980) that subjects searching for a given target letter make most letter-detection errors on the most frequent function words. Subjects in this experiment made more errors on the function words than on the content words, which were less frequent in English.

Thus, the complete pattern of results cannot be explained solely in terms of Corcoran's (1966) hypothesis that subjects tend to scan the acoustic image of the target word in the course of the letter-detection task. The simple notion of phonetic encoding during silent reading fails to account for the higher error percentages observed with function than with content words. Corcoran's (1966) explanation for the high error rates on the word *the*, which were more than double those on words containing silent *es*, was that the word *the* is a highly redundant word, which may be taken for granted and thus not scanned. The present results demonstrate, first, that the same missing-letter effect holds for other, less frequent, and presumably less redundant function words (mean frequency 1,854 as opposed to 69,971 for *the*), and second, that it holds even for the scrambled-word passage, in which the occurrence of any of the test words cannot be predicted on the basis of the preceding word context. Furthermore, the present results demonstrate that the difference in error percentages between pronounced and silent *es* is, if anything, much greater for the function words than for the content words, which is contrary to what one might expect if the function words were indeed redundant and therefore not scanned.

The pattern of results obtained in standard-case and mixed-case passages is also more consistent with our model than with Corcoran's (1966) phonetic recoding hypothesis. In our view, subjects make most letter-detection errors on the frequent function words in prose and scrambled-word passages because they tend to process highly frequent words in terms of units larger than the letter. The use of mixed-case passages impedes the formation of such reading units and might be expected, in effect, to unpack the processing of function words, making their constituent letters more visible. Consequently, error rates on function words and, to a lesser extent, on content words should be lower for the prose mixed-case passage relative to the prose standard-case passage, as was indeed observed. However, it could be argued that any text manipulation that slows down the reader would make the letters of function words easier to detect. Our earlier data (e.g., Drewnowski & Healy, 1977; Healy, 1976) suggest, though, that only manipulations causing a spatial-configural disruption have this effect. The use of nonsense scrambled-word passages instead of prose slows down the reader but does not alter the relative proportion of errors on the word *the* (Healy, 1976). Another possible reason for fewer errors on the mixed-case passage is that capital letters may be easier to find than lowercase letters. Yet, even if such a result were obtained, it could not explain the selective drop in errors for frequent function words, which was not seen for content words.

Finally, the present data indicate that the observed silent-*e* effect may be due in large part to the differential location of the target letter *e* within the test word. Subjects searching the scrambled-letter passage for instances of the letter *e* made significantly more errors on the terminal ("silent") locations than on the intermediate ("pronounced") locations within the letter strings. This finding points to the presence of a strong target-location effect (which was in fact observed by Corcoran) and suggests the need for another experiment in which the target-letter location within the word is rigidly controlled.

EXPERIMENT 2

In Experiment 2, we controlled for letter location by insuring that all target letters, both silent and pronounced, occurred in the penultimate position in the unstressed final syllable of a test word. Because of this constraint, the present comparison was between silent *es* and reduced or schwa-type *es*, rather than between silent *es* and nonreduced or full *es*. However, the schwa-type *e* is in fact a very frequent realization of *e* and, hence, presumably qualifies as a modal (typically pronounced) *e* (cf. Locke, 1978). Furthermore, the phonetic form of *e* (/i/, /ɛ/, or /ə/) was found by Corcoran (1966) to have no influence on the frequency of letter-detection errors.

To control for the linguistic class of the test words, only content words were used. We also controlled two additional variables that were reported to affect the rate of letter-detection errors: (1) the length and frequency of the words containing the target letter (Drewnowski & Healy, 1977; Healy, 1976, 1980), and (2) the linguistic environment of the target letter, which occurred either in a morpheme suffix or in the word stem (Drewnowski & Healy, 1980). Finally, we employed a passage typed with standard typecase as well as one with mixed typecase, as we had in Experiment 1, to determine how voicing and the other variables tested interact with visual factors.

Method

Subjects. Ninety-six Yale undergraduates participated as subjects. The first 28 of them received course credit for their participation; the remaining 68 were paid $1.00 each.

Design and materials. Two 240-word nonsense passages were constructed. The passages included 48 test words, each of which included a single instance of the letter *e* in the penultimate position. The test words were classified into eight groups of six words, on the basis of three orthogonal divisions: (1) words with *e* as part of a terminal morpheme suffix (e.g., *higher*) vs. words with *e* as part of the stem (*order*); (2) words in which the *e* is pronounced (*higher*) vs. words in which the *e* is silent (*worked*)[2]; (3) short words (1-2 syllables) of high frequency (mean = 220; range 101-605; Kučera & Francis, 1967) (*higher*) vs. long words (3 syllables) of low frequency (mean = 6, range 1-12) (*container*). Word length and frequency were treated here as a single variable, since longer English words are typically less frequent than shorter ones.

The specific test words employed are listed in Table 3. Note that three of the six words with a pronounced *e* in the suffix end in *-er* and three end in *-ed* for both the long infrequent and the short frequent words. This division allowed us to make two more controlled comparisons: The first was an assessment of test word ending (suffix vs. stem), including only words ending in *-er*. The second was an assessment of the effects of voicing, including only words ending in the *-ed* suffix. For these comparisons, the terminal letters in the word (*r* or *d*) were not confounded with any of the critical variables.

The passages also included 48 foil words matched as closely as possible in syllabic length and frequency to the 48 test words (so that a subject could not determine whether a word contained a target on the basis of length or frequency alone), 48 filler words in the frequency range of 11-12, 48 filler words in the frequency range of 114-148, and 48 function filler words with frequency greater than or equal to 461. None of the foil or filler words included the letter *e*, except for one filler word (*stopped*), which was included erroneously and was therefore not included in the error analyses reported below.

The test, foil, and filler words were arranged in the passage at random, with the constraint that every block of five successive words include one test, one foil, and three filler words, one of which was a function word. No punctuation was included in the passage except for a final period.

Table 3

Test Words Used in Experiment 2

	Test Word Ending			
	Suffix		Stem	
Voicing	Common	Rare	Common	Rare
Pronounced	higher	container	order	wallpaper
	longer	blackmailer	summer	midsummer
	lower	narrower	mother	hamburger
	added	contracted	system	nitrogen
	started	disgusted	market	unravel
	wanted	discarded	women	caramel
Silent	worked	diminished	sides	syllables
	walked	commissioned	times	microphones
	passed	malnourished	values	disclosures
	turned	impassioned	rates	limousines
	asked	uniformed	sales	contributes
	showed	abolished	states	signatures

The two passages differed only in terms of letter capitalization. The standard-case passage was typed with only the initial letter of the initial word capitalized. The mixed-case passage was prepared in two versions: Even letters were capitalized in the even version and odd letters in the odd version.

Each subject was shown the standard-case passage along with either the even or odd version of the mixed-case passage. Half the subjects were shown the odd version and half were shown the even version. Each passage was typed on a separate sheet of paper and photocopied for distribution to the subjects. The order of presentation of standard- and mixed-case passages was perfectly counterbalanced across subjects. Copies of the two passages preceded by a consent form and a sheet of instructions were stapled together into a booklet for each subject.

Procedure. The procedure was essentially the same as that used in the previous experiment, except that subjects were run in groups of one to six.

Results

The results are summarized in Table 4, which includes for each of the two passage types (standard case and mixed case) the mean error percentages (and standard errors of the mean) as a function of the voicing of the target letter (pronounced vs. silent), the frequency and the length of the test word (common vs. rare), and test word ending (suffix vs. stem).

The subjects made more errors on short (common) words (19.2%) than on long (rare) words (14.3%), $F(1,95) = 27.8$, $p < .01$, and on the standard-case version (22.6%) than on the mixed-case version (11.0%) of the passage, $F(1,95) = 123.5$, $p < .01$. The observed difference in error rates between common and rare words was greater for the standard-case passage (8.2%) than for the mixed-case passage (1.5%). This significant interaction, $F(1,95) = 20.3$, $p < .01$, can be attributed to the fact that processing in the mixed-case passage largely occurs at the letter level.

Neither of the remaining variables yielded the expected effects. First, there was no difference in errors made on targets occurring in word stems (16.9%) and those occurring in word suffixes (16.6%). Second, slightly more errors were made on words in which the target e was pronounced (17.4%) than on words in which the target e was silent (16.1%). This difference was not statistically reliable, $F(1,95) = 2.6$, $p > .10$, but there was a significant interaction between voicing and passage type, $F(1,95) = 27.9$, $p < .01$. More errors were made on silent than on pronounced targets in the mixed-case passage (11.9% vs. 10.0%), but the opposite result was obtained in the standard-case passage (20.4% vs. 24.8%).

A further pair of comparisons was made to determine whether the failure to find the expected effects of voicing and word-ending type (suffix vs. stem) was due to a partial confounding of these factors with the specific terminal letter of the word. In the first analysis, which involved only items in which the target was pronounced and only those ending in -er, words in which the target occurred in the stem were compared with those in which the target

Table 4

Means (and Standard Errors) for Error Percentages as a Function of Passage Type, Voicing of the Target Letter, Frequency of Test Word, and Test Word Ending for Experiment 2

	Passage Type							
	Standard case				Mixed case			
	Suffix		Stem		Suffix		Stem	
Voicing	Common	Rare	Common	Rare	Common	Rare	Common	Rare
Pronounced	29.51 (2.58)	17.88 (2.07)	28.99 (2.42)	22.92 (2.29)	10.42 (1.44)	7.99 (1.35)	9.55 (1.37)	12.15 (1.65)
Silent	25.52 (2.49)	18.40 (2.45)	22.92 (2.52)	14.58 (1.96)	12.67 (1.61)	10.59 (1.62)	14.06 (1.95)	10.24 (1.57)

occurred in the suffix. Even for these words, more errors were made when the target occurred in the stem (24.7%) than when in occurred in the suffix (16.8%), rather than the opposite, $F(1,95) = 24.6$, $p < .01$. In the second analysis, which involved only items in which the target occurred in the suffix -ed (and hence none of those ending in nasal or liquid consonants), words in which the target was pronounced were compared to those, matched in terms of frequency, in which the target was silent. There was no overall difference between errors on silent es (16.8%) and on pronounced es (16.1%), $F(1,95) < 1$, and for the standard-case passage alone, slightly more errors were made on pronounced es (23.6%) than on silent es (21.4%). It is therefore clear that the failure to find the expected effects of voicing and word-ending type cannot be attributed to the specific terminal letter of the word.

Discussion

The primary purpose of this experiment was to determine whether the effects of letter voicing obtained in Experiment 1 could be attributed to the voicing of the target letter or to letter location. When letter location was strictly controlled, the typical effects of voicing--more errors on silent than on pronounced targets--were not obtained. Instead, no overall difference between silent and pronounced letters was found, and, in fact, a small difference in the direction opposite to that predicted was found for the standard version of the passage. Thus, the effects of letter voicing in Experiment 1 may be due to the confounding of voicing and letter location. Although Corcoran (1966) did control for letter location in one of his data analyses and still obtained significant effects of voicing, he did not control for word class or word frequency. It is possible that these factors may have influenced his results: For example, words with terminal or penultimate silent es may have included a disproportionate number of common words (e.g., are, have, or used). Furthermore, Corcoran's sample of pronounced es most likely included some that were stressed (e.g., he, be, or met), whereas our sample did not.

Not only does the present study fail to demonstrate the expected effects of voicing, but it also fails to demonstrate the expected effects of word ending: No more errors were made on letters occurring in word suffixes than on those occurring in word stems. This result is not in agreement with our earlier report (Drewnowski & Healy, 1980). However, the earlier study dealt with the suffix -ing, whereas the present study deals with the suffixes -er and -ed. In fact, we previously noted that other suffixes, including -ment, -ion, and -en, did not yield as many detection errors as -ing, and that -ing was special in a number of ways, including its high frequency and high spatial predictability. For that reason, it is not surprising that the morpheme suffixes we used in the present study did not yield a preponderance of detection errors.

In contrast, word frequency and length did yield large effects in the expected direction. In accord with the unitization model, many more errors were made on the short common words than on the longer rare words, and this effect was greatly diminished when every other letter was typed in capital letters.

EXPERIMENT 3

After controlling for target location and word frequency in Experiment 2, we failed to observe phonetic effects in the letter-detection task. However, all *es* were unstressed, and we were dealing exclusively with phonetic attributes at the letter level. Perhaps phonetic factors play a larger role at some higher level in the linguistic hierarchy. For example, subjects may make more errors on unstressed than on stressed syllables, since the stressed syllables would be expected to be more salient in a phonetically recoded version of the text. Therefore, in Experiment 3, we selected test words in which the target letter *e* either did or did not carry the primary word stress. In addition, we used both relatively frequent and relatively infrequent test words. We expected frequent words to be read at the syllable level or above and rare words to be read letter by letter. Consequently, the effects of syllable stress should be greater for the more frequent words.

As in previous experiments, we used standard-case and mixed-case passages. Since the formation of reading units larger than the letter should be impeded in the mixed-case passage, the effect of stress should be greatly reduced by means of this purely visual manipulation. In addition, because the results of Experiment 1 attest to the importance of target-letter location within the word, we now used three-syllable test words with the target letter occurring in the first, second, or third syllable of the word.

Method

Subjects. Ninety-six students at the University of Toronto served as volunteer subjects in this experiment, conducted in a classroom setting.

Design and materials. Two 240-word scrambled-word passages were constructed for the present experiment. Each passage included the same 48 test words, with each word containing a single instance of the target letter *e*. The test words were classified into 12 groups of words on the basis of three orthogonal divisions: (1) the location of the target letter *e*, which was in the first, second, or third syllable of the word (e.g., *certainly*, *attention*, *incorrect*); (2) the presence or absence of primary word stress on the syllable containing the target letter (e.g., *certainly* vs. *decision*), and (3) the frequency in the language of the test word (e.g., *certainly* vs. *decimal*). The mean frequency of the more common words was 99.9 (Kučera & Francis, 1967) and the mean frequency of the less common words was 6.6. The high-frequency test words stressed on the third syllable were necessarily less common than the remaining words in the high-frequency category. The specific test words employed are listed in Table 5. Note that the linguistic structure of the test words is not constant. For example, many test words with *es* in the first and second syllables end in morpheme suffixes (e.g., *certainly*), but those with *es* in the last syllable mostly do not. For that reason, we cannot be certain at this point that we have successfully controlled for the potential effects of other linguistic variables.

The two passages were composed of the 48 test words, 48 foil words selected to match the test words in number of syllables and approximate frequency, 96 filler words selected from an article in *Psychology Today*, and

Table 5

Test Words Used in Experiment 3

Target Location

Syllable 1		Syllable 2		Syllable 3	
Stressed	Unstressed	Stressed	Unstressed	Stressed	Unstressed
		High Frequency			
certainly	decision	attention	suddenly	incorrect	consider
regular	religion	directly	properly	discontent	character
technical	beginning	successful	governor	introspect	apartment
medical	specific	professor	powerful	indirect	citizens
		Low Frequency			
decimal	mechanic	collector	prophecy	circumvent	immodest
terminal	revision	pathetic	prosperous	dispossess	disorders
democrats	permitting	compelling	tolerant	misdirect	concurrent
sensory	semantic	appendix	numbering	unconcern	transparent

Table 6

Means (and Standard Errors) for Error Percentages as a Function of Passage Type, Word Frequency, and Syllable Stress in Experiment 3

Passage Type

	Standard Case		Mixed Case	
Frequency	Stressed	Unstressed	Stressed	Unstressed
High	12.7 (1.6)	18.3 (1.8)	12.6 (1.5)	14.2 (1.6)
Low	9.7 (1.5)	9.7 (1.3)	10.6 (1.7)	10.5 (1.5)

48 of the most common function words selected from Kučera and Francis (1967). The foil words, filler words, and function words did not contain any instances of the letter *e*: All instances of *e* in the passage thus occurred in the 48 test words.

The sequence of words in each passage was constructed with the same constraints used in Experiment 2. The two passages--standard case and mixed-case (odd and even versions)--were presented to the subjects in a counterbalanced order. Instructions to subjects and details of the testing procedure were the same as described in the previous experiments.

Results

The results are summarized in Table 6, which includes for each of the two passages the mean error percentages (and standard errors of the mean) as a function of test word frequency and the presence or absence of stress on the letter *e*.

The subjects made more errors on the high-frequency than on the low-frequency test words, $F(1,95) = 39.6$, $p < .01$, in agreement with previous results. Overall, more errors were made on unstressed than on stressed *es*, $F(1,95) = 8.6$, $p < .01$, but the effect of stress was only found for the high-frequency words. The significant interaction between word frequency and stress, $F(1,95) = 9.4$, $p < .01$, is consistent with an earlier report (Smith & Groat, 1979) and supports our hypothesis that common words are more likely to be read in syllable-size units and that, therefore, phonetic effects are more likely to occur at the syllable level. Further support for this hypothesis was provided by the finding that the effects of frequency and the effects of stress were larger in the standard-case passage, in which units larger than the letter could be formed, than in the mixed-case passage, in which the formation of such reading units was impeded. The interaction of frequency and passage type was significant, $F(1,95) = 5.4$, $p < .05$, and there was a weak interaction of stress and passage type, $F(1,95) = 3.8$, $p = .05$.

The effects of target location and syllable stress are shown in Figure 1 separately for standard-case and mixed-case passages. Target location (first, second, or third syllable of the test word) significantly affected error rates: More errors were made on *es* in the second and third syllables of test words than on *es* in the first syllable, $F(2,190) = 8.6$, $p < .01$. As in Experiment 1, the target-location effect was higher for the standard-case passage than for the mixed-case passage: The interaction of passage type and target location was significant, $F(2,190) = 6.6$, $p < .01$.

These results suggest that subjects use reading units of different sizes at the different locations within the word. If subjects reading three-syllable words do make use of reading units larger than the letter only in the second and third syllables of words, then we might expect the effect of syllable stress to interact with target location; the effect of stress should be greater in the later locations within the test word, especially in the standard-case passage. In accordance with these predictions, we found significant interactions between target location and syllable stress, $F(2,190) = 5.0$, $p < .01$, and between passage type, target location, and syllable stress, $F(2,190) = 5.0$, $p < .01$.

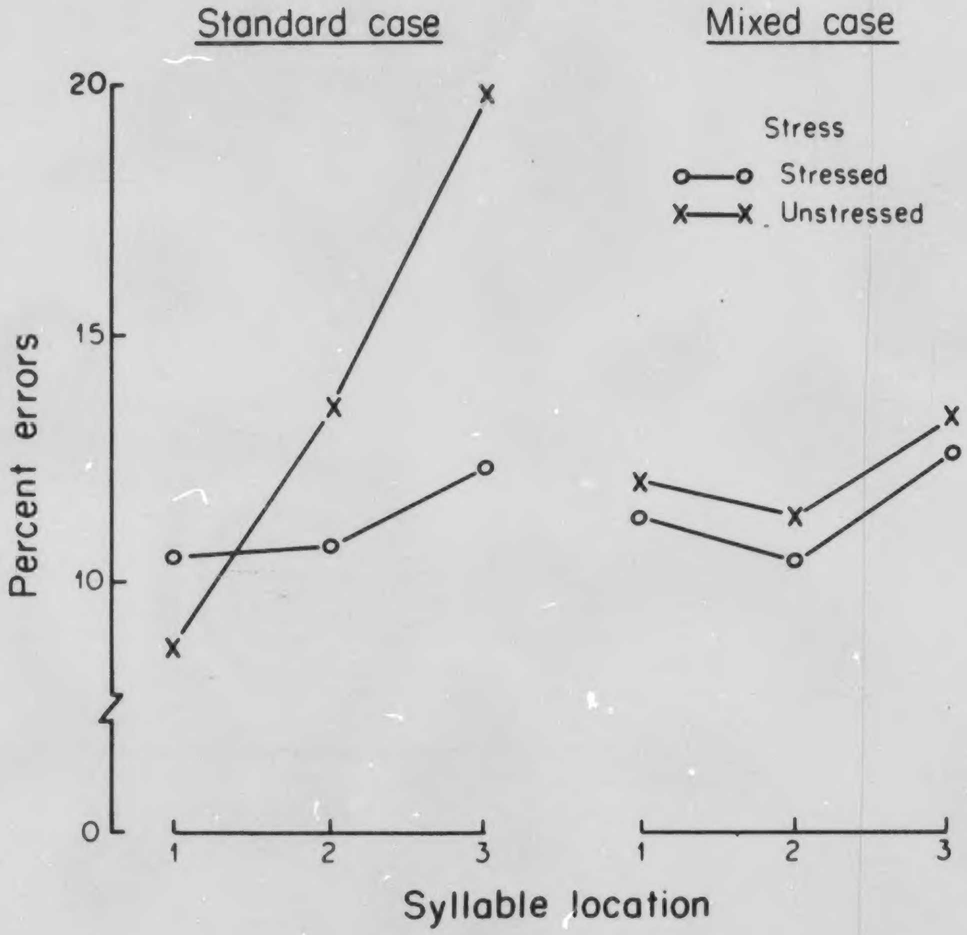


Figure 1. Error percentages as a function of passage type, syllable stress, and target location in Experiment 3.

We have proposed that infrequent words are less likely than frequent words to be read in syllable units. Hence, the effects of passage type, target location, and syllable stress should be more evident for the high-frequency test words. Figure 2 shows error percentages as a function of word frequency, target location, and syllable stress. Only the data from the standard-case passage are included, since little difference between the various conditions was observed for the mixed-case passage. The highest error scores were obtained for unstressed *es* occurring in the second and third syllables of high-frequency test words. There was a significant interaction between word frequency and target location, $F(2,190) = 6.3$, $p < .01$, and a significant four-way interaction among word frequency, passage type, target location, and syllable stress, $F(2,190) = 3.5$, $p < .01$. (Note, however, that the observed drop in error rate for third-syllable stressed *es* in high-frequency words may be partly due to the fact that these words were relatively infrequent, as noted in the Method section.)

Discussion

In Experiment 3, we found significant effects of syllable stress, with more errors made on targets occurring in unstressed than in stressed syllables. However, these effects were by no means general but, rather, occurred only under narrowly defined circumstances. Effects of stress were not observed for the mixed-case passage, for infrequent test words, or for targets occurring in the first syllable of test words. We propose a common explanation for the lack of stress effects in each of these cases: Because we assume that the stress effect is a phonetic effect at the syllable level, effects of stress should be absent when no units larger than the letter are used. Consequently, the effects of stress should be attenuated for mixed-case passages and for infrequent words. We also propose that subjects read longer content words in terms of different-size units at different locations of the word. Specifically, subjects may process the first syllable of three-syllable words to the point of identifying each letter but use reading units larger than the letter in later locations of the word. By this explanation, the effects of stress, which occur at the syllable level, are most likely to be found in the later locations of relatively common words, as was indeed observed. To summarize, it appears that stress effects occur only when the subject is able to form reading units at the syllable level.

It is possible that the observed difference between *es* in unstressed and stressed syllables is due to a difference between reduced (or schwa-type) and nonreduced (or full) *es*. All stressed *es* in the present experiment were nonreduced, whereas *all* unstressed *es* in *the* second and third syllables (but not the first syllable) were reduced. This explanation is consistent with the observed interaction between target location and syllable stress, but it cannot account for the interactions between syllable stress and test word frequency or between syllable stress and passage typecase.

It can also be argued that the location effect found here (and the similar effect in Experiment 1) is caused by subjects' scanning only the initial syllable of the test word for target letters and failing to scan the remainder of the word. However, this account is not consistent with our finding a stress effect for the second and third syllables of test words. If subjects failed to scan the end of the word, then there should not be a difference between stressed and unstressed syllables at the end of the word.

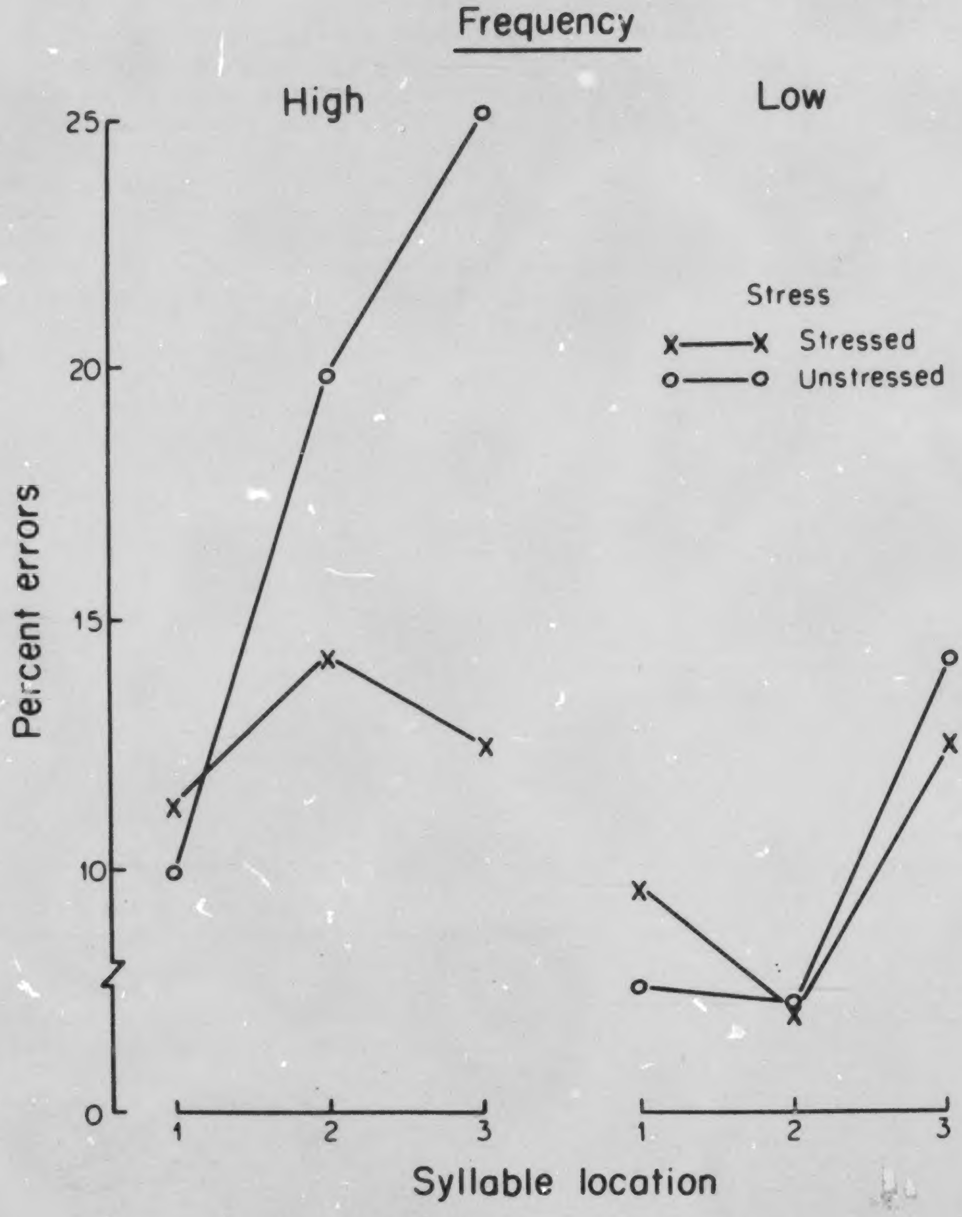


Figure 2. Error percentages as a function of word frequency, syllable stress, and target location for standard-case passage in Experiment 3.

It may seem puzzling that subjects make many errors on short frequent words (e.g., the) and few errors on the initial syllables of longer, less frequent words (e.g., certainly). However, the unitization model is compatible with these results because of the postulate that subjects move their attention to the next word in the text, without completing processing at the letter level, once they have identified a particular configuration as a word. For example, subjects will not complete processing of the letters in the word the once they have identified the familiar configuration as a word.

GENERAL DISCUSSION

One recurring issue in studies of reading is the extent to which phonetic factors are involved in the process of silent reading. Although many studies have addressed this issue (see McCusker, Hillinger, & Bias, 1981, for a detailed review), most were limited to situations involving the presentation of isolated words and pseudowords, as in the lexical decision task. Little is known about the extent of phonetic recoding in the course of normal comprehension of printed text.

Our technique of letter detection in prose contexts (Corcoran, 1966; Healy, 1976) provides a good index of performance during normal silent reading. Indeed, the pattern of errors on the letter-detection task can be used as a reading diagnostic, since we have demonstrated in developmental studies that error rates vary as a function both of the reading materials and of the subjects' reading skill (Drewnowski, 1978, 1981). In our previous studies, we have used the letter-detection technique to examine the size of the units employed in reading printed text. In contrast, most investigators using this technique have focused on the phonetic recoding hypothesis, by comparing error rates either on silent and pronounced letters (Chen, 1976; Coltheart et al., 1975; Corcoran, 1966; Mohan, 1978; Smith & Groat, 1979) or on modal (typically pronounced) and nonmodal (atypical) phonemes (Locke, 1978). However, the use of normal English in this task carries with it important confoundings. In the present study, we designed special passages to eliminate the confoundings of target-letter location, test word frequency, and linguistic context in order to determine whether the voicing of the target letter has a residual effect on error rates.

The study provided further support for the unitization model. All three experiments revealed clear effects of word frequency: More errors were made on frequent than on less frequent words. In Experiment 1, word frequency covaried with linguistic class (function vs. content words), whereas in Experiment 2, word frequency covaried with word length. Both factors were controlled in Experiment 3, which included only three-syllable content words and still revealed a significant effect of word frequency. This frequency effect is consistent with the previous observations by Healy (1976, 1980) and supports the hypothesis that subjects are more likely to read common than rare words in units larger than the letter, even in the case of long content words. The effect of frequency is considerably more dramatic for the most frequent function words the and and (see Drewnowski & Healy, 1977).

In agreement with previous reports (e.g., Corcoran, 1966), we found in Experiment 1 that subjects made more errors on silent than on pronounced es.

However, the voicing of the target letter covaried with letter location within the word. A similar difference between "silent" and "pronounced" locations was found in the scrambled-letter passage composed of unpronounceable letter strings, suggesting that letter location rather than letter voicing might be the more important factor. When the location of the target letter was strictly controlled, as it was in Experiment 2, no effects of voicing were obtained. The effects of voicing noted by previous investigators who did control for location may have been due to a confounding of letter voicing and word frequency. In Experiment 2, no effect of voicing was obtained either for high-frequency or low-frequency test words.

We did obtain phonetic effects in Experiment 3 in which we systematically manipulated syllabic stress, rather than letter voicing. The subjects made more errors on targets occurring in unstressed than in stressed syllables. We interpret these results as indicating that the phonetic representation of text may be scanned at the level of the syllable, rather than at the level of the letter. The observation that the effects of syllable stress and word frequency were greatly diminished in passages in which every other letter was typed in capitals supports the view that both these effects operate at levels above the level of the letter. In addition, the observation that the effect of stress is most evident for the more frequent words supports our hypothesis that such words tend to be processed in syllable-size units.

We also found in Experiment 3 that the effects of word frequency and syllabic stress were most marked for targets in the second and third syllables of test multisyllabic words. These data suggest that the initial syllables of multisyllabic words are processed to the point of letter identification, regardless of the test word frequency or its syllabic stress pattern.

These results are consistent with the basic notion originally put forth by Corcoran (1966) that subjects looking for target letters scan a phonetically recoded version of text during silent reading. However, the phonetic units scanned do not appear to be at the letter level, but rather, at the level of the syllable. Our data suggest that these syllable units are formed only under certain conditions; their formation depends on word frequency, on the location of the syllable within the word, and on the visual features of printed text. The present study thus reconciles two distinct hypotheses regarding the reading process--phonetic recoding and unitization--and places these hypotheses within a single theoretical framework.

REFERENCES

Chen, K. Acoustic image in visual detection for deaf and hearing college students. Journal of General Psychology, 1976, 94, 243-246.

Coltheart, M., Hull, E., & Slater, D. Sex differences in imagery and reading. Nature, 1975, 253, 438-440.

Corcoran, D. W. J. An acoustic factor in letter cancellation. Nature, 1966, 210, 658.

Drewnowski, A. Detection errors on the word the: Evidence for the acquisition of reading levels. Memory & Cognition, 1978, 6, 403-409.

Drewnowski, A. Missing -ing in reading: Developmental changes in reading units. Journal of Experimental Child Psychology, 1981, 31, 154-168.

Drewnowski, A., & Healy, A. F. Detection errors on *the* and *and*: Evidence for reading units larger than the word. *Memory & Cognition*, 1977, *5*, 636-647.

Drewnowski, A., & Healy, A. F. Missing *-ing* in reading: Letter detection errors on word endings. *Journal of Verbal Learning and Verbal Behavior*, 1980, *19*, 247-262.

Foss, D. J., & Blank, M. A. Identifying the speech codes. *Cognitive Psychology*, 1980, *12*, 1-31.

Healy, A. F. Detection errors on the word *the*: Evidence for reading units larger than letters. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, *2*, 235-242.

Healy, A. F. Proofreading errors on the word *the*: New evidence on reading units. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, *6*, 45-57.

Kučera, H., & Francis, W. N. *Computational analysis of present-day American English*. Providence, R.I.: Brown University Press, 1967.

LaBerge, D., & Samuels, S. J. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, *6*, 293-323.

Locke, J. L. Phonemic effects in the silent reading of hearing and deaf children. *Cognition*, 1978, *6*, 175-187.

McCusker, L. X., Hillinger, M. L., & Bias, R. G. Phonological recoding and reading. *Psychological Bulletin*, 1981, *89*, 217-245.

Mohan, P. J. Acoustic factors in letter cancellation: Developmental considerations. *Developmental Psychology*, 1978, *14*, 117-118.

Schindler, R. M. The effect of prose context on visual search for letters. *Memory & Cognition*, 1978, *6*, 124-130.

Smith, P. T., & Groat, A. Spelling patterns, letter cancellation and the processing of text. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language 1*. New York: Plenum Press, 1979.

FOOTNOTES

[1] With the exception of a few short function words (e.g., *he*), words ending in a single terminal pronounced *e* are few and have a low frequency in the language (e.g., *adobe*, *apostrophe*). For that reason, we did not use an orthogonal manipulation of voicing and location. Similarly, there are no content words in English that are comparable in frequency to the common function words, so we did not attempt an orthogonal manipulation of word frequency and word function.

[2] As far as we can tell, our division of words into those with pronounced or silent *es* corresponds to the syllabic/nonsyllabic classification of Smith and Groat (1979). The one exception is the word *values*, which would have been classified by them as a syllabic *e* but was classified by us as a silent *e*.

CATEGORICAL PERCEPTION: ISSUES, METHODS, FINDINGS*

Bruno H. Repp

1. INTRODUCTION

Ever since the beginning of language--and perhaps even earlier--human beings have classified things and events into categories. Categorization occurs when we focus on important properties that are common to different objects and ignore irrelevant detail. Although such an act of attention is commonly accompanied by verbal statements, categorization may also occur covertly. However, the fact that most categories do have names is definitely advantageous in communication. For example, the name of an object or event may still be recalled when memories of physical details have long faded. It is not surprising, therefore, that category names form the core of our vocabulary.

Many of the categories we have are natural--they reflect obvious physical partitions among things in the world, and there is little question or choice as to what is included in a particular category, and what is not. Other categories, however, are less transparent and may reflect special knowledge or conventions. Some scientific categories fall in this class; for example, the zoologist's category of fish excludes dolphins and whales but includes eels and sea horses, whereas a prescientific, shape-oriented category of fish might include the former but exclude the latter. In addition, there are cases, such as those involving aesthetic judgment or preference, where it is up to the individual to draw the boundaries between categories; and categories based on relative judgment (size, weight, speed, etc.) are totally situation-specific and essentially arbitrary.

The categories of speech--which include the phonetic segments or phonemes--play an important part in linguistic theory and are implicated in the development and continued use of alphabetic writing. However, illiterates have little awareness of them (Morais, Cary, Alegria, & Bertelson, 1979); nonlinguists know them only in a vague fashion, commonly mistaking letters for

---

*This review article is to appear in Volume 9 of Speech and Language: Advances in Basic Research and Practice, edited by N. J. Lass (Academic Press, 1983).

Acknowledgment. Preparation of this paper was supported by NICHD Grant HD01994. I am grateful to Robert Crowder, Carol Fowler, Virginia Mann, Ignatius Mattingly, David Pisoni, Michael Studdert-Kennedy and, especially, Sigfrid Soli for helpful comments on an earlier draft, and to Marilyn Parnell for her incomparable keyboard skills.

[HASKINS LABORATORIES: Status Report on Speech Research SR-70 (1982)]

phonemes; and even among specialists there are disputes about their precise nature and description. Did linguists merely invent these categories for the purpose of abstract description, or did they discover an important, though not very transparent, principle of discrete organization that underlies human speech production and perception? And if the latter, do the proposed descriptive categories map directly onto the _functional_ categories of active speech communication? These questions are aspects of the more general question about the _psychological reality_ of the products of linguistic analysis--an issue that lies at the heart of modern psycholinguistics.

Categorical perception research in the speech domain is concerned with the perceptual reality of phonetic segments--that is, with the role of phonetic categories in perceptual processing _regardless_ of whether the perceiver has any awareness of them. Although categorical perception research is in principle a rather broad area of inquiry permitting a variety of methods, it has over the years become identified with a particular laboratory paradigm. That paradigm has generated a large amount of useful research that presents a challenge to theories of speech perception. However, in recent years there have been some signs of exhaustion. This seems a good time to review some of the history, methods, and problems of categorical perception research, and to try to see where we stand. We will begin with a historical overview. The studies mentioned therein will be discussed in greater detail in later sections.

2. HISTORICAL OVERVIEW

2.1. The Early Haskins Research

Categorical perception research began at Haskins Laboratories not long after the construction of the first research-oriented speech synthesizer, the Pattern Playback. Liberman, Harris, Hoffman, and Griffith (1957) used this new tool to construct a series of syllables spanning the three categories /b/, /d/, and /g/ preceding a vowel approximating /e/. Although these stimuli formed a physical continuum (obtained by increasing the onset frequency of the second formant in equal steps), listeners classified them into three rather sharply divided categories. To test whether the physical differences among the stimuli within a category could be detected by listeners, Liberman et al. employed an ABX discrimination task. (This task requires subjects to indicate whether the last of three successive stimuli matches the first or the second, which are always different from each other.) The results showed that stimuli classified as belonging to different categories were easily discriminated, while stimuli perceived as belonging to the same category were very difficult to tell apart, even though the physical differences seemed comparable. This characteristic pattern of results came to be called "categorical perception" (see Section 3.1). By assuming that listeners have no information beyond the phonetic category labels (an assumption later often referred to as the "Haskins model"), Liberman et al. (1957) were able to generate a fair prediction of discrimination performance from known labeling probabilities; however, performance was somewhat better than predicted, suggesting that the subjects did have some additional stimulus information available.

The pioneering experiment of Liberman et al. (1957) set the pattern for a number of similar studies exploring different kinds of phonetic contrasts. Thus, Liberman, Harris, Kinney, and Lane (1961) reported categorical perception of the /d/-/t/ contrast cued by "first-formant cutback"; Liberman, Harris, Eimas, Lisker, and Bastian (1961) found similar results for the intervocalic /b/-/p/ distinction cued by closure duration; and Bastian, Eimas, and Liberman (1961) demonstrated that stop manner cued by closure duration (/slIt/-/splIt/) was likewise categorically perceived. These findings contrasted with those of Fry, Abramson, Eimas, and Liberman (1962) and Eimas (1963), who showed that synthetic vowels forming an /ɪ/-/ɛ/-/æ/ continuum were discriminated equally well within and between phonetic categories--a result referred to as "continuous perception." Continuous perception was obtained also with other properties of vowels such as duration (Bastian & Abramson, 1964) and intonation contour (Abramson, 1961), as well as with nonspeech stimuli that had certain critical features in common with categorically perceived speech stimuli (e.g., Liberman, Harris, Eimas, Lisker, & Bastian, 1961; Liberman, Harris, Kinney, & Lane, 1961). Thus, categorical perception seemed to be specific to speech (excluding isolated vowels), and to stop consonants in particular.

These early findings provided one of the pillars for the motor theory of speech perception set forth by the Haskins group (Liberman, 1957; Liberman, Cooper, Harris, MacNeilage, & Studdert-Kennedy, 1967; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). The basic tenet of the motor theory is that speech perception and articulatory control involve the same (or closely linked) neurological processes. When different phonetic categories are distinguished by essentially discrete articulatory gestures (as with stop consonants differing in voicing or place of articulation), perception of stimuli from a physical continuum spanning these categories will be categorical; on the other hand, when continuous articulatory variations between phonetic categories are possible (as with the vowels), perception will be continuous (cf. Liberman, Harris, Eimas, Lisker, & Bastian, 1961, p. 177). In other words, the motor theory takes categorical perception to be a direct reflection of articulatory organization.

For a number of years, categorical perception research stayed at Haskins Laboratories--a situation that changed only in the 1970s when appropriate speech synthesizers became available in other laboratories. The only pertinent research outside Haskins in the early years was conducted by Harlan Lane and his collaborators at the University of Michigan, who examined categorical perception from a psychophysical viewpoint, focusing on the question whether a similar phenomenon could be produced with nonspeech stimuli under comparable experimental conditions. The results of that not very successful effort were summarized in Lane's (1965) critical review of the early Haskins research. Lane's criticisms anticipated some of the concerns of later researchers, but they had little impact at the time because they were backed up by rather weak data. However, they provoked a forceful, if somewhat belated reply by Studdert-Kennedy, Liberman, Harris, and Cooper (1970), which remains the classic statement of the Haskins view of categorical perception (see Section 3.1).

Categorical perception research continued at Haskins during the 1960s. Abramson and Lisker (1970) showed that the voiced-voiceless distinction for utterance-initial stop consonants, as cued by voice onset time, was categorically perceived by speakers of two languages with different voicing boundaries, Thai and English. Another early cross-language study was conducted by Stevens, Liberman, Öhman, and Studdert-Kennedy (1969) with Swedish and English vowels. Although perception of these vowels was not quite as continuous as in the earlier study by Fry et al. (1962), there seemed to be no connection between identification and discrimination, suggesting noncategorical perception. The categorical perception of the place-of-articulation distinction for voiced stop consonants (Liberman et al., 1957) was replicated by several studies, including one by Mattingly, Liberman, Syrdal, and Halwes (1971), who, for the first time, included stop consonants in utterance-final position, as well as several nonspeech controls that were not categorically perceived.

2.2. The Information Processing Approach

In the meantime, two Japanese scientists became interested in the Haskins findings and began to experiment along similar lines. The work of Fujisaki and Kawashima (1968, 1969, 1970, 1971), presented in a series of limited-circulation progress reports, remained virtually unknown in the West until Pisoni (1971, 1973, 1975) discussed and extended it. The work of these authors, of Pisoni in particular, brought categorical perception into the mainstream of contemporary psychology. While, up to this time, the focus had been on categorical perception as a pure *phenomenon*, on its relation to articulatory behavior, and on the effects of learning on auditory sensitivity, attention now turned to perceptual *processes* and to stimulus and task variables involved in categorical perception experiments.

Fujisaki and Kawashima (1969, 1970, 1971) formulated a dual-process model for the discrimination of speech stimuli, which explicitly distinguished between categorical phonemic judgments and judgments based on auditory memory for acoustic stimulus attributes (see Section 3.2). Thus, the model attempted to account for the commonly observed difference between the categorical predictions of the Haskins model and actual discrimination performance--a difference that was treated as an uninteresting nuisance in the early Haskins research (unless it was sufficiently large to be interpreted as "continuous" perception). Fujisaki and Kawashima also explored new classes of speech stimuli (synthetic fricatives, semivowels, and liquids) and showed that their perception was somewhat less categorical than that of stop consonants, though not as continuous as that of isolated vowels. They further experimented with vowels of varying duration, with or without added context, and showed that even vowels may be perceived quite categorically when conditions are unfavorable for auditory memory. The imaginative (though somewhat fragmentary) work of Fujisaki and Kawashima has served as a stimulus for further research to the present day (see Sections 4.1 and 5.1).

Several ideas of the Japanese researchers were elaborated and tested by Pisoni (1971, 1973, 1975; Pisoni & Lazarus, 1974), who applied the dual-process model to a variety of discrimination paradigms, showing that the categoricalness of perception depends, to some extent, on how much use can be made of auditory memory in a task. He further confirmed this point by varying stimulus duration, the duration of interstimulus intervals, and by introducing interfering sounds between the stimuli to be discriminated. Pisoni and Tash

(1974) were the first to use same-different reaction times as an indicator of subjects' sensitivity to acoustic stimulus differences within phonetic categories. This analytic research began a trend of increasing interest in subjects' ability to discriminate subphonemic (within-category) acoustic differences between speech stimuli--a trend that shifted the emphasis from categorical perception as a mere phenomenon to the psychoacoustics and psychophysical methodology of speech discrimination.

2.3. Offsprings of Categorical Perception Research

The early 1970s spawned several significant research developments that grew out of categorical perception research and have since become highly active areas semi-independent from (but, of course, intimately related to) the traditional approach to categorical perception, with which they share the use of the classic experimental paradigm requiring identification and discrimination of synthetic speech sounds from a physical continuum. The diversification proceeded on three fronts--new subjects, new tasks, and new stimuli.

One of the new enterprises was research on infant speech perception. In a now classic paper, Eimas, Siqueland, Jusczyk, and Vigorito (1971) reported that 1- and 4-month-old human infants responded to stimuli from a voice-onset-time (/ba/-/pa/) continuum in a way similar to adults: The infants discriminated stimuli from opposite sides of the adult category boundary (as indicated by an increase in the rate of non-nutritive sucking in response to a stimulus change), but not physically different stimuli from the same category. This exciting finding has since been replicated several times and has been extended to a variety of different stimuli. Infant speech perception research has been following closely on the heels of the research on adult speech perception, and, on the whole, it has revealed that infants' perceptual capabilities are remarkably similar to those of adults, though without the influence of specific linguistic experience. Important research is now under way to determine the role played by exposure to a specific language in the course of perceptual development (see Section 6.3).

A second development concerns studies of animal speech perception. Although few in number, they have attracted much attention through Kuhl and Miller's (1975, 1978) finding that chinchillas divide a voice onset time continuum into the same categories as adult humans do. There is increasing activity today in this methodologically difficult but fascinating area (see Section 6.4).

On the methodological side, researchers began to experiment with a variety of discrimination paradigms and different response measures, including rating scales, reaction time, and even evoked potentials (see Section 4.2). The phenomenon of categorical perception held up remarkably well under this onslaught. A vigorous strand of research was started by Eimas and Corbit (1973), who applied the technique of selective adaptation to continua of synthetic speech stimuli. By presenting one or the other endpoint stimulus over and over, it was possible to shift the location of the phonetic category boundary, and even to shift the associated discrimination peak with it. Numerous studies, including some of the most elegant work in speech perception, have tried to unravel the sources and mechanisms of the adaptive shifts. Unfortunately, the returns have been somewhat disappointing, for it is now quite clear that the adaptation effect does not take place at the level of "phonetic feature detectors," as originally believed, but is a purely auditory phenomenon (Roberts & Summerfield, 1981; Sawusch & Jusczyk, 1981). While the selective adaptation technique continues to be useful for probing into the auditory processes of speech perception, this research is tangential to the concerns of this review and will not be discussed in detail. (For reviews, see Ades, 1976; Cooper, 1975; Diehl, 1981; Eimas & Miller, 1978.)

Categorical perception research also continued along more traditional lines with adult human subjects. Encouraged by the increasing sophistication of speech synthesis, however, researchers explored phonetic categories other than those of stop consonants and vowels. More or less categorical perception was demonstrated for the affricate-fricative distinction (Cutting & Rosner, 1974), for continua of liquid consonants (McGovern & Strange, 1977; Miyawaki, Strange, Verbrugge, Liberman, Jenkins, & Fujimura, 1975), of nasal consonants (Larkey, Wald, & Strange, 1978; Miller & Eimas, 1977), and of the oral-nasal distinction (Miller & Eimas, 1977), among others. With certain qualifications, this research showed that virtually all consonantal distinctions are categorically perceived (see Section 5.2).

2.4. The Psychophysical Approach

In the early Haskins research and in Lane's (1965) critical review of it, a good deal of attention was paid to the possibility that categorical perception was caused by general auditory processes. The conclusion from the early Haskins studies (notwithstanding Lane's objections, which had only weak empirical support) had been that categorical perception was specific to speech, and to (stop) consonants in particular. Interest in the psychoacoustics of categorical perception reawakened in the mid-1970s, when the earlier conclusion was shattered by several demonstrations of apparently categorical perception of nonspeech sounds. Thus, Cutting and Rosner (1974) claimed to have found categorical perception of complex tones varying in rise time (the "pluck"-"bow" distinction); Miller, Wier, Pastore, Kelly, & Dooling (1976) reported categorical perception of noise-buzz sequences intended to be analogous to a voice-onset-time continuum; and Pisoni (1977) found similar results for two tones varying in relative onset time. In Section 5.3, we will examine these and other studies in considerable detail.

The demonstrations of categorical perception of nonspeech sounds stimulated some psychophysicists to take a closer look at categorical perception, and some speech researchers to take a closer look at psychophysics. Thus, Macmillan, Kaplan, and Creelman (1977) attempted to fit categorical perception into the framework of signal detection theory; Ades (1977) made a cautious (and still largely unexplored) connection with the related psychophysical work of Durlach and Braida (1969; Braida & Durlach, 1972); Pastore (1981) reviewed psychoacoustic factors that may be relevant to categorical perception; and Schouten (1980) went so far as to propose that all of speech perception could be explained by psychoacoustic principles.

Psychophysical theories were further encouraged by several reports of successful speech discrimination training. While earlier studies had focused on the role of learning in categorical perception and had attempted (with limited success) to produce the phenomenon by training subjects in the use of category labels for nonspeech stimuli (e.g., Cross, Lane, & Sheppard, 1965; Parks, Wall, & Bastian, 1969), Carney, Widin, and Viemeister (1977), for example, took the converse approach: They showed that categorical perception of speech may be attenuated by training listeners to pay attention to acoustic stimulus properties. These findings suggested that categorical perception is essentially a function of experience and attentional strategies (see Section 6.1).

Underlying these psychophysical approaches is a single-process (or "common-factor") view of categorical perception, which assumes that linguistic categories are essentially psychoacoustic in nature (Miller et al., 1976; Pastore, Ahroon, Baffuto, Friedman, Puleo, & Fink, 1977). This view has emerged in recent years as a serious competitor for the dual-process model proposed by Fujisaki and Kawashima (see Section 3.4). The antagonism between these two models has become tied up with the more general controversy about whether it is necessary to postulate a special phonetic mode of perception at all (cf. Liberman, 1982; Repp, in press; Schouten, 1980).

The psychophysical trend stimulated researchers at Haskins Laboratories and elsewhere to illustrate the complexity of phonetic perception in new experiments. The emphasis of much of this new research is on the complex, many-to-one relationship between acoustic stimulus properties and phonetic percept, demonstrated experimentally as phonetic "trading relations" or other contextual interactions between several different acoustic cues. Since many of these studies use the methodology of categorical-perception research (i.e., identification and discrimination of stimuli from synthetic speech continua), they may be viewed as dealing with the categorical perception of stimuli varying along two or more dimensions (e.g., Best, Morrongiello, & Robson, 1981; Fitch, Halwes, Erickson, & Liberman, 1980), with particular attention to the distinction between auditory and phonetic modes of perception. This research has led to various contemporary versions of the motor theory (e.g., Bailey & Summerfield, 1980; Repp, Liberman, Eccardt, & Pesetsky, 1978). Several recent studies have been particularly successful in constructing appropriate nonspeech analogs to examine the presumed speech-specificity of the demonstrated cue trading relations (Best et al., 1981; Summerfield, in press). We will discuss some of these studies below; for detailed reviews, however, see Liberman (1982) and Repp (in press).

Investigators have also shown an increased interest in one aspect of the methodology of categorical perception--contextual dependencies among successive stimuli in a labeling or discrimination task (Crowder, 1982; Healy & Repp, 1982; Repp, Healy, & Crowder, 1979; see Section 3.3). Related work has grown out of the research on selective adaptation (Diehl, Elman, & McCusker, 1978; Sawusch & Nusbaum, 1979). This is likely to be an area of considerable activity in the near future.

We have come to the end of this brief historical review, in the course of which I hope to have mentioned all major trends and landmarks. In the following, more detailed review, I focus in sequence on the several different factors that contribute to the phenomenon called "categorical perception." Discussions of theoretical and methodological issues (Sections 3 & 7) precede and follow the core sections (4, 5, & 6), which are dedicated to the review of data.

3. EMPIRICAL ASSESSMENT OF CATEGORICAL PERCEPTION: MODELS AND METHODS

3.1. Defining Categorical Perception: The Classical Haskins View

The preceding section has provided a broad answer to the question of what constitutes categorical perception. Now we shall examine this issue in somewhat more detail. First, it is useful to point out that the term "categorical" may be understood in at least three different ways, which may be called "literal," "phenomenal," and "empirical."

Literally speaking, categorical perception refers to the use of categories by an individual in responding to his or her environment. In this sense, it is a ubiquitous phenomenon not restricted to speech, and in particular there is no implication that the perceiver is unaware of stimulus variations within a category. This is not the way in which the term has been used by speech researchers, but others have occasionally interpreted and used it that way.

Phenomenally speaking, categorical perception refers to the experience of discontinuity as a continuously changing series of stimuli crosses a category boundary, together with the absence of clearly perceived changes within a category. It must be emphasized here that categorical perception is a very striking and readily demonstrated phenomenon. Anyone who sits down and listens to one of the standard series of stop consonants varying in voice onset time or formant transitions, provided he or she is able to hear the synthetic sounds as speech, will experience abrupt perceptual changes at certain places on the continuum. The continuing attraction of categorical perception to both the novice and the seasoned investigator lies in its permanent and replicable vividness in the listener's experience.

However, subjective experience alone is not enough to satisfy the rigors of scientific investigation, and we must therefore turn to categorical perception as an empirical concept, describing a particular pattern of data in an experiment. It is here that the situation becomes more complex, because ideal categorical perception (where category labels are the sole determinant of performance) is rarely, if ever, encountered in the laboratory. Empirical data typically deviate more or less from this ideal, and some criterion must be applied for deciding whether they do or do not provide evidence for categorical perception. In fact, to capture different amounts of deviation, it may be necessary to speak of degrees of categorical perception (cf. Studdert-Kennedy et al., 1970, p. 238), although this violates the strict definition of categorical perception proposed by the Haskins group:

> "Categorical perception refers to a mode by which stimuli are responded to, and can only be responded to, in absolute terms. Successive stimuli drawn from a physical continuum are not perceived as forming a continuum, but as members of discrete categories. They are identified absolutely, that is, independently of the context in which they occur. Subjects asked to discriminate between pairs of such 'categorical' stimuli are able to discriminate between stimuli drawn from different categories, but not between stimuli drawn from the same category. In other words, discrimination is limited by

identification: subjects can only discriminate between stimuli that they identify differently" (Studdert-Kennedy et al., 1970, p. 234, their emphasis).

A typical experiment might proceed as follows: In an *identification (labeling) test*, stimuli from a physical continuum, spanning two categories unambiguously represented by the endpoint stimuli, are presented repeatedly in randomized order to subjects for classification into one or the other category. In a subsequent (sometimes preceding) *discrimination test*, typically using the ABX paradigm, adjacent or more widely separated stimuli from the continuum are presented for discrimination. The identification data are summarized in the form of *labeling functions*, which relate response percentages to stimulus location on the continuum. The discrimination data yield one or more *discrimination functions*, which relate a measure of discrimination accuracy (usually percent correct) for stimulus pairs of equivalent physical separation to stimulus location. Ideal categorical perception in this standard design exhibits four semi-independent characteristics:

(1) Labeling probabilities change abruptly somewhere along the continuum; in other words, the identification functions have a rather steep slope. The point of maximum slope is the *category boundary* (equivalently defined as the point at which responses in two adjacent categories are equiprobable).
(2) Discrimination functions show a peak at the category boundary; that is, stimuli are more easily discriminated when they fall on opposite sides of the boundary than when they fall on the same side.
(3) Discrimination performance within each category is at or near chance level.
(4) Discrimination functions are perfectly predictable from the labeling probabilities (using one of the simple formulae provided by the Haskins model--see Pollack & Pisoni, 1971). This implies that (a) the discrimination peak is in exactly the right place and of the right height, and (b) the labeling probabilities are appropriate, i.e., they apply independently of the context in which they were observed. (These two corollaries show that criterion 4 is not directly implied by criteria 1, 2, and 3.)

As we have already observed, the actual data are rarely perfect. They may fit the ideal description more or less well. In evaluating the data, more importance is attached to some criteria than to others. For example, the criterion of steepness of labeling functions is a very weak one. Given that stimulus continua do contain ambiguous stimuli in the category boundary region, the steepness of labeling functions depends in part on how closely the stimuli are spaced along the continuum. (See the discussion of this issue by Lane, 1965, and by Studdert-Kennedy et al., 1970.) A much more important criterion is the presence of a peak in the discrimination function that coincides with the location of the phoneme boundary--a feature of the data later christened the *phoneme boundary effect* (Wood, 1976a). It is the essential defining characteristic of categorical perception, although it may not be sufficient if the other criteria are grossly violated. A certain amount of deviation is usually tolerated for both of the remaining criteria (near-chance performance within categories and match of predicted and obtained discrimination functions).

A statistical criterion of whether some data do or do not represent categorical perception is provided by the goodness of fit of the predictions (cf. Healy & Repp, 1982; Pisoni, 1971). In practical usage, however, the striking contrast between the results for stop consonants and isolated vowels (or nonspeech stimuli) has often supported the "categorical-continuous" dichotomy irrespective of any deviations from the ideal patterns of categorical or continuous perception. Later research, however, has yielded a number of intermediate cases that can no longer be accurately characterized by this simple dichotomy.

The question of what constitutes admissible evidence for categorical perception was discussed in detail by Studdert-Kennedy et al. (1970) in their reply to Lane's (1965) critical review. Lane had focused on criterion 1 (described above) and had revealed its weakness, and he had criticized criterion 4 on the basis that corollary 4b may not be satisfied (see Section 3.3 for further discussion of his arguments). Although the Haskins authors were remarkably effective in rebutting Lane's methodological objections, there remained one prime weakness in their presentation. It stemmed, in large measure, from viewing categorical perception as a monolithic phenomenon, and from a resulting unwillingness to consider in detail the different factors that enter the experimental situation defining categorical perception. In a perceptive commentary, Haggard (1970) noted that "the controversy between Lane and the Haskins group stems from a failure to enumerate levels or aspects of the perceptual process and make separate statements about them" (p. 6).

3.2. Speech Perception as a Two-Component Process: The Dual-Process Model

Speech perception was conceived by the Haskins group of the 1950s and 60s as a modular process that, for a given phonetic distinction, is either categorical or continuous. The origin of the two types of phonetic perception was hypothesized to lie in the articulatory continuity or discontinuity of the segmental distinctions perceived; that is, in whether articulations intermediate between those typical of two segments occur in natural speech (or are anatomically possible at all). *Both* types of phonetic perception were thought to be mediated by an articulatory representation of the input, in accord with the motor theory, although the similarity of continuous speech perception and nonspeech perception was evident.

This essentially unidimensional view of speech perception contrasts with the dual-process model introduced by Fujisaki and Kawashima (1969, 1970) and elaborated by Pisoni (1971, 1973, 1975). Rather than assuming that only a single perceptual mode is active at any given time, they proposed that two modes are active simultaneously (or in rapid sequence). One of them is strictly categorical and represents phonetic classification and the associated verbal short-term memory. The other mode is completely continuous and represents processes common to all auditory perception, including auditory short-term memory. The results of any particular speech discrimination experiment are assumed to reflect a mixture of both component processes: The part of performance that can be predicted from labeling probabilities (using the Haskins model) is attributed to categorical judgments, while the remainder (the deviation from ideal categorical perception) is assigned to memory for acoustic stimulus properties.

The dual-process model partially abandons the articulatory rationale for categorical perception by explicitly equating continuous with auditory (i.e., nonspeech) perception. Accordingly, the difference in categoricalness between, say, stop consonants and vowels is hypothesized to derive not from the different articulatory properties of these segments but from the different strengths of their representations in auditory memory. By augmenting the Haskins prediction model with a free parameter representing the contribution of auditory memory, Fujisaki and Kawashima also introduced a way of quantifying different degrees of categorical perception that, unfortunately, has not been adopted by other researchers.

It is obvious that the dual-process model opened up new avenues for research. It now became possible to ask how subjects in an experiment utilize the two sources of information (categorical and continuous, or: phonetic and auditory), and what factors might lead them to rely more on one than on the other. Since the continuous component was identified with general auditory memory, several standard experimental techniques became available to weaken or strengthen that memory and to observe the subsequent changes in speech discrimination performance. Attention turned from categorical perception as a somewhat mysterious, "special" speech phenomenon to an analysis of the experimental situation--of the task factors, stimulus factors, and subject factors that conspire to generate a particular pattern of results.

3.3. Problems of Prediction: Context Effects versus Phonetic Mediation

At this point, a brief digression into the methodology of predicting discrimination performance is in order, since the prediction test is the most widely used formal criterion of categorical perception. The Haskins model derives its predictions of perfectly categorical discrimination from labeling probabilities obtained in an independent identification task in which the individual stimuli are presented in random order (see Pollack & Pisoni, 1971, for computational techniques). This procedure was criticized by Lane (1965) on two grounds. First, he argued, the phonetic categories assumed to be employed covertly in the discrimination task may not be identical with the ones employed overtly in the labeling task. Second, even if the same categories were used, the probabilities of classifying the stimuli into the different categories may not be the same in the two tasks because the labeling probabilities may be sensitive to context (i.e., they may be influenced by immediately preceding or following stimuli), and the context of individual stimuli is different in the two tasks. Of course, these arguments applied only to cases of apparently noncategorical perception; they reflected Lane's contention that categorical perception was not specific to speech and could be acquired in the laboratory (see Section 5.3).

The first objection is the less serious of the two. For many continua of speech sounds, there are no plausible alternative phonetic categories to the ones intended and suggested to the subjects by the experimenter. In other cases, the objection may be valid but could be met by not restricting the subjects' response set in the labeling task. However, although individual differences in the number and kind of categories used may come to the fore in a free-response situation, subjects are also rather willing to adopt categories suggested by the experimenter, even if they are not the standard ones (see Carden, Levitt, Jusczyk, & Walley, 1981, for a recent striking example).

Therefore, it seems that a mismatch of phonetic categories in identification and discrimination tasks has not been a serious problem in categorical perception research. (A related, but more subtle, problem that cannot be so easily dismissed is that subjects may devise phonetic subcategories in a discrimination task, based on different degrees of confidence in their phonetic judgments--e.g., "good /b/" vs. "poor /b/"; see Liberman, Harris, Eimas, Lisker, & Bastian, 1961, for an early documented example. We will encounter this issue again later in this review.)

The second objection, that of context effects in labeling, deserves closer attention. Studdert-Kennedy et al. (1970) responded to it by insisting that "categorical perception entails context-free perception" (p. 246). In other words, if context effects are present and lead to a mismatch of predicted and obtained discrimination performance, that is simply evidence that perception is not categorical. Lane (1965) suggested that the predictions be derived by having subjects label the stimuli in the same context in which they are presented for discrimination. (For early applications of this method, see Cross & Lane, 1964--cited in Lane, 1965--and also Fujisaki & Kawashima, 1969.) However, Studdert-Kennedy et al. (1970) dismissed this procedure on the grounds that "by 'acknowledging context,' we predict discrimination from discrimination" (p. 247).

This response is characteristic of the unidimensional view of categorical perception espoused by the Haskins group at that time. Their sole concern was to determine whether or not perception of a given set of stimuli was categorical. Although they acknowledged that ideal categorical perception is rarely encountered, they were not particularly interested in the causes of the deviations from the ideal. However, an explanation of these deviations is likely to increase our understanding of categorical perception, particularly since there are many instances of "noncategorical" perception that are far from "continuous." It is possible to distinguish three such situations (Healy & Repp, 1982): (1) There may be context effects in (covert) phonetic labeling, but the subjects may nevertheless rely exclusively on category labels in discriminating different stimuli. (This is certainly a form of categorical perception, though not the absolute one of the Haskins definition.) (2) Labeling may be independent of context, but subjects may utilize auditory stimulus information in discrimination and thereby exceed the predictions of the Haskins model. (In this case, perception is absolute without being categorical.) (3) The deviations from the categorical ideal may be due to both contextual effects in labeling and auditory memory in discrimination.

These considerations suggest that phonetic mediation (reliance on category labels) in discrimination and context sensitivity in labeling are two logically distinct aspects of the experimental situation that can (and should) be assessed separately. To assess phonetic mediation, the predictions of discrimination performance are derived from "in-context" labeling probabilities, i.e., from subjects' labeling responses to stimuli presented in the exact sequence used also in the discrimination task; any remaining discrepancies between predicted and obtained performance may then be unambiguously attributed to auditory memory. The magnitude of context effects in labeling, on the other hand, may be inferred directly from the "in-context" labeling responses by examining contextual contingencies (Fujisaki & Kawashima, 1969; Healy & Repp, 1982; Repp et al., 1979).

The separation of context sensitivity and phonetic mediation is essentially an elaboration of the dual-processing hypothesis. It provides more realistic estimates of labeling probabilities and, thereby, a more accurate assessment of the relative contributions of (covert) categorical judgments and auditory memory to discrimination. Indeed, it appears that the small advantage of obtained over predicted discrimination scores, which is customarily obtained with stop consonants, may be entirely due to contrast effects in (covert) labeling, and not to any direct access to auditory memory (Healy & Repp, 1982). Context effects may themselves have a dual-process explanation: They may either represent a form of response bias at the level of phonetic categorization (see, e.g., Diehl et al., 1978; Shigeno & Fujisaki, 1980), or they may derive from an interaction of auditory memory traces akin to lateral inhibition (Crowder, 1978, 1981), or both factors may be at work simultaneously.

3.4. Psychoacoustics and Categorical Perception: The Common-Factor Model

The dual-process hypothesis of Fujisaki and Kawashima contains the assumption that categorical perception derives entirely from the phonetic component in the model, i.e., from the application of linguistic categories. The auditory component is assumed to be essentially continuous. There is an alternative possibility, however: It could be that some auditory dimensions of speech are not continuous, and that there are psychoacoustic thresholds that may coincide with the phonetic category boundaries on a speech continuum. In other words, categorical perception may be a phenomenon of auditory perception, in part or in toto. Pastore et al. (1977) introduced the term common-factor model for the hypothesis that "a single (common) factor [other than phonetic categorization--BHR] causes both a peak in the discrimination function and a categorical dichotomy and thus the correlation between the two" (p. 686). This proposal was encouraged by the early findings of seemingly categorical speech discrimination in human infants (Eimas et al., 1971), and in nonhuman animals (Kuhl & Miller, 1975), and of certain nonspeech stimuli by human adults (Cutting & Rosner, 1974; Miller et al., 1976), and it has come to play a central role in contemporary speech perception research. It is so important because it promises not only to explain the speech perception capabilities of infants and animals, but also to provide a principled account of the demarcation and evolution of linguistic categories.

According to the common-factor model, the discrimination peak that characterizes categorical perception (the "phoneme boundary effect") comes about because, given a psychoacoustic threshold on a continuum, different subthreshold stimuli are mutually indiscriminable, sub- and suprathreshold stimuli are easy to tell apart, and different suprathreshold stimuli are discriminated according to Weber's law, which predicts increasingly poorer performance as stimulus differences of constant absolute size move away from the threshold (cf. Miller et al., 1976). The difficulty with the common-factor model does not lie in its proposal that discrimination peaks can come about in this way (for they obviously can, as several studies of nonspeech continua have shown--see Section 5.3) but in the difficulty of showing that they do have a strictly psychoacoustic basis in the case of speech continua that are categorically perceived.

To obtain support for this hypothesis, some authors have employed signal detection theory or related methods to derive the "perceptual spacing" of stimuli on a speech continuum, characteristically finding that stimuli are spaced further apart in the boundary region than within categories (Elman, 1979; Macmillan et al., 1977; Oden & Massaro, 1978; Perey & Pisoni, 1978). However, this result merely amounts to a re-description of the data; it does not answer the question of why stimuli are spaced in this way in perception. As we will see in later sections, the various attempts at proving that specific auditory thresholds underly particular phonetic boundaries have not been uniformly successful, although some have produced encouraging results.

Another problem for the common-factor model is that there are cases of "boundary effects" on continua that quite clearly do not straddle any psychoacoustic thresholds. These include continua of isolated vowels (e.g., Pisoni, 1971), isolated fricative noises (Fujisaki & Kawashima, 1970), or musical intervals (e.g., Burns & Ward, 1978). The results of these studies suggest (as does some of the research reviewed in Section 6) that a discrimination peak may be caused simply by the existence of appropriate categories. On the other hand, we do have some rather strong evidence for psychoacoustic discontinuities on certain speech continua (see Pastore, 1981). Perhaps, what is needed is a modified dual-process model--one that admits the possibility of significant nonlinearities in auditory perception while, at the same time, assuming a separate contribution of phonetic category labels in the process of discrimination.

This modified dual-process model might be considered unparsimonious by some, but it does appear to accommodate the existing evidence, as the following review will attempt to show. The model also bears a certain resemblance to the two-factor model of Durlach and Braida (1969; Braida & Durlach, 1972), although their model was developed to account for discrimination of sound intensity (a true psychoacoustic continuum over most of its range). The Durlach-Braida model assumes two components, a "sensory-trace mode" and a "context-coding mode," which jointly contribute to discrimination accuracy and differ in their relative permanence. The relevance of this model to categorical perception was pointed out by Ades (1977). If two processes are necessary to account for simple intensity resolution, it can hardly be unparsimonious to postulate two separate processes in speech perception.

It can be seen from the foregoing discussion that theoretical reasoning in categorical perception research has not progressed very far. The models proposed so far are simple and few in number. They contrast with the richness and occasional complexity of the data, to which we now turn. The following three sections are dedicated to a review of research on categorical perception within the confines of the standard identification-discrimination paradigm. Some relevant research using unconventional methods will be mentioned in the concluding section. The organization of the three sections is based on the view that categorical perception, as a pattern of experimental results, is a joint function of three major factors: task variables, stimulus variables, and subject variables. Categorical perception is not a property attached to a particular stimulus set. Rather, it is a way in which a particular individual responds to particular stimuli in a particular experimental situation. Accordingly, Sections 4-6 divide the evidence into pieces relating to task, stimulus, and subject factors. Although it would be logical to begin with the most important section (that on stimulus factors), it seemed more convenient to treat task factors first, in order to avoid prolonged discussions of methodology in the following sections.

4. TASK FACTORS IN CATEGORICAL PERCEPTION

In this section, we will examine to what extent categorical perception is a function of the task used to assess discrimination. There are two ways of pursuing that question: Either one starts with stimuli that are not very categorically perceived (e.g., isolated vowels) and tries to make their perception more categorical by modifying the task; or, conversely, one starts with stimuli whose perception is highly categorical and attempts to make their perception less categorical. Both approaches have been used in the past. Within the framework of the dual-process model, they amount to either decreasing or increasing the auditory memory component in subjects' performance. The contribution of the categorical component is assumed to be either constant or inversely proportional to that of auditory memory.

4.1. Procedures for Increasing Categorical Perception

There are two ways of reducing auditory memory without changing the stimuli themselves or their relationship. (See Section 5.1 for effects of stimulus manipulations.) One is to introduce interference in the form of noise or by interpolating irrelevant sounds between the stimuli to be discriminated. The other way is to increase the temporal separation of the stimuli, so that auditory memory for the first stimulus has decayed by the time the second stimulus arrives.

4.1.1. Interference With Auditory Memory

In the earliest vowel discrimination study, Fry et al. (1962) found no discrimination peaks at category boundaries, but this was probably due to a ceiling effect, coupled with the use of imperfectly controlled stimuli. Most later studies (e.g., Fujisaki & Kawashima, 1969, 1970; Pisoni, 1971; Stevens et al., 1969) have found fairly clear peaks on vowel continua, so there is good reason to believe that there is a phonetic component in vowel discrimination. Cross and Lane (1964; cited in Lane, 1965) actually used the original tapes of Fry et al. and added noise in the form of an additional, irrelevant resonance. Although it seems that phonetic identification should have suffered considerably, Lane (1965) nevertheless reports that marked discrimination peaks were observed at the category boundaries.

Fujisaki and Kawashima (1969, 1970) included a condition in which a constant /a/ vowel immediately followed each of the test stimuli (vowels from an /i/-/e/ continuum, presented in ABX triads for identification and discrimination). They claimed to have found more nearly categorical perception in that condition than when the fixed context was omitted, and they attributed that difference to the context serving as a "perceptual reference." By this they presumably meant that it facilitated categorization and also, perhaps, that it interfered with auditory memory. Their data are less than clear, however, and this is compounded by the fact that different data are reported in their 1969 and 1970 papers for ostensibly the same experiment. The 1970 data, in particular, show a narrowing of the discrimination peak coupled with an increase in within-category discrimination performance. Thus, the context did not seem to interfere with auditory memory, although it may have aided categorization.

Fujisaki and Kawashima also reported that adding a constant vocalic context to fricative noise stimuli from a /ʃ/-/s/ continuum had little effect on discrimination performance (which, curiously, was highly categorical even for isolated fricative noises), although closer inspection of their results again reveals that within-category discrimination was improved by the presence of context. These results contrast with recent data that suggest that a following vowel reduces the discriminability of fricative noises, even in subjects who are able to perceptually segregate the noise from the vowel (Repp, 1981c), and that isolated noises are not categorically perceived (Healy & Repp, 1982; Repp, 1981c).

Pisoni (1975, Exp. III) examined the role of a fixed context in more detail. He argued that, if the context stimuli serve as a perceptual anchor, as hypothesized by Fujisaki and Kawashima, then it should not matter whether the context precedes or follows the test stimuli. If, on the other hand, the context interferes with auditory memory, one might expect that a following context will produce more interference than a preceding one. In addition, Pisoni hypothesized that the similarity of context and test stimuli would determine the amount of interference. To test this last hypothesis, Pisoni used four different sounds (a 1000-Hz pure tone, a burst of white noise, and the vowels /a/or /ɛ/) as contexts for stimuli from an /i/-/I/ continuum. The context immediately preceded or followed each test stimulus in labeling and ABX discrimination tests, with a no-context control condition included. The results supported the similarity hypothesis: Discrimination scores were lowest in the /ɛ/-vowel context, although all contexts lowered performance somewhat. There was also more of a decrement when the context followed, rather than preceded, the test stimuli, although the difference was small.

Pisoni made no attempt to assess the degree of categorical perception in the various context conditions, nor did he report whether labeling probabilities were influenced by the various contexts. To examine these issues, Repp et al. (1979) presented pairs of vowels from an /i/-/I/-/ɛ/ continuum in a same-different discrimination task. The interval between the two stimuli on a trial was either silent or partially filled by an irrelevant vowel sound (/y/). The intervening stimulus produced a clear decrement in discrimination performance, and a comparison with predictions from standard identification data led to the conclusion that perception had become more categorical. However, Repp et al. also had their subjects label the stimuli in pairs and computed "in-context" predictions of discrimination performance (see Section 3.3). These predictions matched the obtained scores much better than did the standard predictions and, significantly, the match was equally good whether or not an interfering sound was present, even though discrimination scores (as well as the predictions) were much lower in the presence of interference. Evidently, the interpolated sound affected both in-context labeling and discrimination. The effect on labeling was evident in a drastic reduction of contrast effects between the members of a stimulus pair (i.e., of the tendency to assign them different labels).

These results permit two interpretations. The one preferred by Repp et al. (1979; see also Crowder, 1981) was that auditory memory had its effect before phonetic categorization in the form of contrastive interactions between auditory stimulus traces, and that discrimination was subsequently based in large part on phonetic labels, even though the stimuli were isolated vowels. To account for the remaining difference between predicted and obtained discrimination performance (which was considered negligible by Repp et al., but turned out to be rather large in a later, similar study by Healy & Repp, 1982), it seems necessary to appeal either to the covert use of additional phonetic categories in discrimination or to some more permanent form of auditory memory that is immune to interference (such as Massaro's, 1975, "synthesized auditory memory"). The other interpretation is that labeling and discrimination were both based directly on auditory stimulus representations, so that interference with auditory memory affected both equally. In this view, which is congenial to psychophysical theories and seems more parsimonious, labeling is viewed simply as a form of coarse-grained discrimination, and contrast effects in labeling are the consequence, not the cause, of accurate discrimination. However, the presence of peaks in the discrimination function indicated that phonetic categories did influence the subjects' "same-different" decisions at some stage.

Whichever interpretation is preferred, the Repp et al. (1979) data clearly demonstrated that interference with auditory memory has a large effect in a categorical perception task. They are also consistent with the research on the so-called suffix effect--the increase in recall errors for the last item in a word list when that list is followed by another, irrelevant item (Crowder, 1971, 1973a, 1973b; Crowder & Morton, 1969). The traditional interpretation of this effect has been that the suffix disrupts a precategorical auditory trace lasting a few seconds--a trace that retains primarily vocalic information because of its higher distinctiveness (Crowder, 1971; Darwin & Baddeley, 1974). Vowel discrimination tasks probably tap the same kind of memory.

4.1.2. Decay of Auditory Memory

Let us now turn to studies that attempted to manipulate auditory memory by changing the temporal interval (interstimulus interval = ISI) between stimuli to be discriminated. In the context of categorical perception research, this method was first applied by Pisoni (1971, 1973), who introduced variable ISIs (0-2 sec) in a same-different discrimination task using both vowels (/i/-/I/) and stop consonants (/bæ/-/dæ/, /ba/-/pa/). There was a clear decrement in vowel discrimination performance as the interval increased (except for reduced scores at the zero interval), whereas there was little effect on stop consonant discrimination performance. A breakdown of the data into within-category and between-category discrimination scores revealed that both scores decreased for vowels, whereas only a slight decrease in between-category performance could be seen for stop consonants. (Within-category discrimination of stop consonants was close to chance.) Very similar results were obtained in a replication by Cutting, Rosner, and Foard (1976) and, in related studies, by Cowan and Morse (1979) and Repp et al. (1979) for vowels, and by Frazier (1976) for consonants.

Since between-category discrimination of vowels was thought to be based on category labels, Pisoni concluded from the uniform decline in performance that an increase in temporal delay resulted in a decay not only of auditory memory (of which there was very little for stop consonants) but also of phonetic memory. However, it seems unlikely that phonetic short-term memory for a single label would decay at all over 2 sec (cf. Fujisaki & Kawashima, 1971). Therefore, all decrements observed were probably due to auditory memory decay.

One question not answered by these studies is whether the memory decay has any asymptote. (Performance continued to decline up to 2 sec.) The question of the time course of memory decay for vowel stimuli was investigated by Crowder (1982a), who varied the ISI in pairs of vowels in a same-different discrimination task, covering the range from 0-5 sec. He found that performance declined up to about 3 sec and then remained stable. In a second experiment of his, the subjects' task was not to respond "same" or "different" but instead to identify the second vowel in each pair. The result was similar: The contextual (contrastive) influence of the first vowel on the second, assumed to be mediated by auditory memory, went away at about 3 sec of separation. (However, see Fujisaki & Shigeno, 1979, for a contradictory finding.) Crowder's results converge with those from suffix effect experiments, where a similar decay rate of auditory memory has been found (Crowder, 1969; however, see Watkins & Todres, 1980). The hypothesis that suffix effects and vowel discrimination are mediated by the same memory store was further supported in a recent study by Crowder (1982b) where he showed that individual differences in the magnitude of the suffix effect correlated reliably with the same subjects' vowel discrimination performance when the interstimulus intervals were short (500 msec) but not when they were long (3 sec).

In summary, these studies leave little doubt that auditory memory plays a role in vowel discrimination tasks, and the parallelism with the suffix effect results suggests that the auditory memory store employed for isolated vowels may also be functional in other tasks involving more complex speech stimuli. The same auditory memory also appears to be responsible for contrastive influences of one stimulus on identification of a following stimulus. (Note, however, that there is also retroactive contrast.) One question that is still not resolved is whether vowel discrimination at delays beyond 3 sec is based entirely on phonetic labels, or whether there is another, more permanent form of auditory memory that aids discrimination at longer delays. Crowder's (1982a) data indicated that the decline in vowel discrimination performance as a function of temporal delay was relatively small while, at the same time, contrast effects in vowel labeling disappeared completely. This suggests that, even at the longest intervals, obtained discrimination performance probably exceeded the in-context predictions (which Crowder did not calculate). Crowder's results appear consistent with the above-mentioned data of Repp et al. (1979), which showed that contrast effects nearly disappeared at a long (filled) interval while obtained discrimination scores were still higher than predicted.

Thus, an explanation of vowel discrimination may ultimately require a three-process model, including two kinds of auditory memory--a fast-decaying one of the kind discussed by Crowder, which mediates contrast effects, and a slower-decaying one that may be utilized in discrimination. The latter corresponds to the "context-coding mode" of Durlach and Braida (1969), and to the "synthesized auditory memory" of Massaro (1975).

The third process, of course, is phonetic categorization. This process is needed in the model to account for the phoneme boundary effects in vowel discrimination, for they could hardly be caused by psychoacoustic thresholds. However, it is possible that these effects, like those on true nonspeech continua (Kopp & Livermore, 1973) and unlike those on stop consonant continua (Elman, 1979; Popper, 1972; Wood, 1976a, 1976b), are entirely due to response bias and not to increased perceptual sensitivity at category boundaries. In other words, there may be no direct "phonetic mediation" in vowel discrimination; rather, the phonetic labels may merely bias auditory judgments. In view of the relative auditory salience of vowel differences, this would not be surprising. One might think of auditory and phonetic decisions being engaged in a race, with auditory decisions winning when the stimuli are isolated vowels but losing when the stimuli are stop consonants. Thus, the influence of phonetic categorization on vowel discrimination may occur by hindsight, as it were, while it may be truly mediational in consonant discrimination.

4.2. Procedures for Reducing Categorical Perception

We turn now to a review of studies that approached the problem of auditory memory from the other side: Instead of reducing discrimination performance (and increasing categorical perception) by decreasing auditory memory, these studies attempted to increase performance (and thereby decrease categorical perception), either by enhancing the auditory memory component or by providing the subjects with finer-grained scales on which to respond. These efforts concentrated on a class of speech sounds that, in the standard experimental setting, were highly categorically perceived and showed little evidence of auditory memory: stop consonants differing in voicing (voice onset time) or place of articulation (formant transitions).

4.2.1. More Sensitive Discrimination Paradigms

Early studies of categorical perception had suggested that stop consonants might not have any representation in auditory memory at all. Although discrimination performance was usually somewhat higher than predicted by the Haskins model, the difference was relatively small and tended to be ignored. Stop consonants were regarded by the Haskins group as abstract perceptual categories stripped of all auditory information, and as the prime example of "encoded" speech sounds whose perception requires the operation of a special speech processor (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman, Mattingly, & Turvey, 1972). Therefore, a demonstration of the existence of some memory for acoustic properties of stop consonants would have been an important contribution.

The ABX discrimination paradigm was used in all early categorical perception studies and remains popular to this day. This paradigm was preferred because it requires a forced choice and, at the same time, absolves the experimenter from specifying the dimension on which the stimuli differ (which, in the case of speech, may be difficult to convey to naive subjects). However, it has often been suggested that ABX is not the most sensitive paradigm, the reason cited being the presumed necessity to compare A and X, with the resulting demand on memory (e.g., Harris, 1952; Pisoni, 1971). Pisoni (1971) tried out a different procedure, the 4IAX paradigm, which shares with the simpler AX (same-different) task the advantage of using pairs rather than triads of stimuli, and with the ABX task the advantage of requiring a forced choice. (In the 4IAX task, the subject must decide which of two stimulus pairs contains a difference.) In Experiment IV of his dissertation, Pisoni found that discrimination of steady-state vowels was improved considerably in the 4IAX paradigm, compared to the ABX paradigm. In his Experiment V, he compared stop consonants from a place-of-articulation (/bæ/-/dæ/-/gæ/) continuum in the same two tasks. Performance in the 4IAX paradigm was only slightly better than in the ABX paradigm, and then only for 2-step comparisons but not for 1-step comparisons. These data did not offer very striking support for an auditory memory component in stop consonant discrimination, although both AXB and 4IAX scores differed reliably from the Haskins model predictions.

In another study using the same two paradigms, Pisoni and Lazarus (1974) examined stop consonants from a voice onset time (/ba/-/pa/) continuum. This study also included a condition in which the subjects were not given the standard labeling test but received instead the /ba/-/pa/ continuum repeatedly in fixed order before doing the discrimination test. This procedure was expected to sensitize the listeners to acoustic stimulus differences. Indeed, there was some increase in performance due to both the 4IAX procedure and the prior experience with the stimulus continuum. However, prior experience appears to have been the critical factor, for Pisoni and Glanzman (1974) failed to find any difference between the ABX and 4IAX paradigms when no pretraining was provided. It should also be noted that in these experiments the difference between the two paradigms was confounded with differences in interstimulus intervals: In the ABX paradigm, there was a 1-sec interval between stimuli in a triad, while in the 4IAX paradigm, the stimuli within a pair were separated by only 150 or 250 msec, with a 1-sec interval between the two stimulus pairs that constituted one trial. The small size of the difference between the two paradigms is consistent with the finding (Pisoni, 1971, 1973) that temporal separation has little effect on stop consonant discrimination.

A direct comparison of the ABX and AX paradigms with speech stimuli was performed recently by Crowder (1982b), who used vowels from an /i/-/I/ continuum and computed d' indices according to the tables published by Kaplan, Macmillan, and Creelman (1978), which make a fair comparison between the two tasks possible. Crowder also made the interstimulus intervals in the two tasks comparable by having the same short (500 msec) or long (3 sec) delays between the B and X items of the ABX triads and between the A and X items of the AX pairs. (The A-B interval in ABX triads was fixed at 250 msec.) The results showed not only that the AX paradigm was more sensitive than the ABX paradigm, but also that it yielded much more stable results, as measured by split-half reliability indices. In Crowder's words, "this result does suggest some caution for investigators choosing the ABX task lest they be making it hard for themselves to demonstrate experimental effects in a sensitive way" (p. 481).

Suspicions that the ABX paradigm encourages categorical perception had been around for some time, and researchers increasingly used alternative paradigms, including oddity (which probably shares all the disadvantages of ABX), AXB (essentially an economical version of 4IAX), 4IAX, and AX. MacKain, Best, and Strange (in press) compared the AXB and oddity paradigms using an /r/-/l/ continuum and found AXB to be superior. A comparison of more than two paradigms for speech discrimination in a single study still remains to be done. However, an extensive comparison of different paradigms for nonspeech discrimination (pure tone frequency or phase relationships) was conducted by Creelman and Macmillan (1979). In contrast to the results with speech, they found greater sensitivity to frequency differences in the (variable-standard) ABX task than in the AX task, with 4IAX performance in between. (However, no differences at all were found between the three paradigms when the task was phase discrimination, suggesting that stimulus factors may interact with task factors in determining discrimination performance.) Another result of the Creelman and Macmillan study was that fixed-standard paradigms (in which only the X stimulus varies from trial to trial) are superior to variable-standard paradigms. Fixed-standard tasks have not been used in speech perception research until fairly recently; since they were usually employed in conjunction with discrimination training, we will review these studies in a later section (6.1).

We should note that it is not quite clear why certain discrimination paradigms are superior to others. Psychophysical theory predicts certain differences for ideal observers (Creelman & Macmillan, 1979), but real subjects are typically far from this ideal. To give a psychological explanation of performance differences, we need a model of the perceptual strategies employed in different tasks, especially in the more complex ones. An unpublished study by Pastore, Friedman, and Baffuto (1976) was directly concerned with that issue. Pastore et al. found for intensity discrimination, as did Creelman and Macmillan for frequency discrimination, that ABX was superior to AX, and that fixed-standard tasks were superior to variable-standard tasks. What is of interest here is that Pastore et al. examined different models of subject strategies in the ABX task and found that the results were best explained by the assumption that only B and X were compared, with A merely serving to "reduce uncertainty." Thus, the data of Pastore et al. do not support the assumption commonly made by speech researchers that listeners compare A and X as well as B and X. However, both sides may be right. The subjects in speech experiments are typically inexperienced, while those in psychophysical experiments are highly practiced. Therefore, it should not be surprising that the latter subjects adopt a more effective strategy. Unless subject strategies also depend on whether the stimuli are speech or nonspeech (as indeed they may), the results available suggest that the ABX paradigm is inferior to the AX paradigm with naive subjects but not with experienced subjects. In Section 6.1, we will discuss the effects of discrimination training on categorical perception. Without such training, it appears that the perception of stop consonants remains fairly categorical, even when more sensitive discrimination paradigms are used.

4.2.2. Rating Scales and Reaction Times

Several researchers have attempted to obtain evidence for subjects' sensitivity to subphonemic detail by modifying the single-item identification task so as to permit the subjects to transmit more information about perceived stimulus differences. One of the earliest studies in that vein was published by Barclay (1972). He presented listeners with a /bæ/-/dæ/-/gæ/ continuum but permitted only two labels, "b" and "g." If subjects' perception had been truly categorical, all stimuli perceived as "d" (as determined in a separate test) should have been assigned to the "b" or "g" categories on a random basis. However, listeners were found to be more likely to apply the label "b" to the more "b"-like instances of /dæ/, and the label "g" to the more "g"-like instances. Thus, listeners showed some sensitivity to acoustic stimulus properties in the center of the continuum. Barclay proposed that categorical perception is primarily a memory phenomenon, observed only when successive stimuli are to be compared. However, Haggard (1970) pointed out that Barclay's stimuli lacked a third formant, which may have created considerable ambiguity in the /dæ/ region. If the intended /dæ/ tokens could indeed be *heard* as either /bæ/ or /gæ/, Barclay's results would seem trivial.

An alternative approach is to provide subjects with a numerical scale on which to rate the individual stimuli. The possibility that categorical perception is merely a consequence of the limited number of phonetic categories available to the perceiver was first investigated by Conway and Haggard (1971; see also Haggard, Summerfield, & Roberts, 1981), who gave their subjects a 9-point rating scale to judge stimuli from 5-member /bIl/-/pIl/ and /gIl/-/kIl/ (voice onset time) continua. The functions relating average stimulus ratings to position on the continuum were distinctly sigmoid in shape, with the largest change in ratings occurring across the phoneme boundary, and virtually no change within categories. If perception had been continuous, the functions should have been linearly increasing. Thus, these results not only provided strong evidence for categorical perception but also offered no indication that a more fine-grained response scale enabled listeners to make distinctions within phonemic categories. In a second, similar study, Conway and Haggard (1971) obtained more continuous-looking functions, but the stimuli spanned only a small range in the vicinity of the boundary, where even the two-category labeling function is nearly linear. Therefore, these data were consistent with categorical perception.

The rating scale of Conway and Haggard had no special relation to the stimuli on the continuum and may have been used by the subjects merely to indicate their degree of confidence in their categorical judgments (as noted by Haggard et al., 1981). Since the endpoints of the scale were explicity identified with phonetic categories, it is perhaps not surprising that categorical perception was obtained. An alternative method is to establish a one-to-one correspondence between stimuli and responses--the task called *absolute identification*. This task was employed by Sachs (1969), whose subjects used the numbers 1-8 to identify eight stimuli from a /bɑdəl/-/bædəl/ continuum, as well as eight stimuli from two /a/-/æ/ continua with different stimulus durations. Despite the procedure used, and despite the fact that the distinction was located in the vowel, perception of the word continuum was quite categorical and so was, to some extent, the perception of the short-duration vowels. (See Section 5.1 for a discussion of effects of phonetic context and duration on vowel discrimination.) These results provided strong evidence that absolute identification does not prevent or even attenuate categorical perception. Later, Cooper, Ebert, and Cole (1976) had their subjects use a 7-point scale to identify stimuli from 7-member /ba/-/wa/ and

/ga/-/ja/ (formant transition duration) continua. Once again, the average numerical responses changed most rapidly across the phoneme boundary, and there was no indication that stimuli strictly within a category (which really applied only to the /ba/ end of the /ba/-/wa/ continuum) were distinguished by the subjects.

Using the same procedure, Perey and Pisoni (1978) compared absolute identification of stimuli from /ba/-/pa/ and /i/-/I/ continua. Once again, the stop consonant data showed categorical perception, while the vowel ratings were more nearly continuous, though not a strictly linear function of stimulus number. Perey and Pisoni showed, however, that stop consonant (and vowel) discrimination in a subsequent ABX test could be predicted more accurately from the rating data than from simple binary labeling probabilities, suggesting that some subphonemic differences were picked up by subjects in the rating task. Still, perception of stop consonants was far from continuous.

Rating scales or absolute identification have been used in many other studies, all of which obtained the basic phenomenon of categorical perception of stop consonants (e.g., Elman, 1979; McNabb, 1976b; Rosen, 1979; Sawusch, 1976). Another variant, the method of direct magnitude scaling, was employed by Port and Yeni-Komshian (1971; cited in Strange, 1972) and Strange (1972). Strange's subjects responded to individual stimuli (stop consonants from a voice-onset-time continuum) by positioning a pointer within a bounded interval. Still, perception remained categorical unless a fair amount of training was provided, in which case some subjects responded more nearly continuously (see Section 6.1).

Yet another approach was recently taken by Samuel (1982). His intention was to locate, for each listener, the "best /ga/" on a narrowly-spaced /ga/-/ka/ (VOT) continuum, presupposing that subjects would be able to distinguish between different stimuli within the /ga/ category. The subjects in this study could control stimulus presentation, step repeatedly through the continuum and zero in on the preferred stimulus. Although Samuel did not determine the reliability of his subjects' estimates of the prototypical /ga/, he did find individual differences that correlated with the magnitude of boundary shifts obtained in a subsequent selective-adaptation experiment. However, since prototype location correlated neither with the location of the phoneme boundary nor with prototype estimates derived by several other procedures (Samuel, 1979), the results must be viewed with some caution.

Studdert-Kennedy, Liberman, and Stevens (1963) found that labeling reaction times for stimuli from stop-consonant and vowel continua exhibited a peak at the category boundary--a finding that has often been replicated (e.g., Pisoni & Tash, 1974; Repp, 1975, 1981a; however, see Hanson, 1977) and is also obtained with nonspeech continua (Cross et al., 1965). Since reaction times indicate the subjects' uncertainty in making phonetic decisions, they are long for ambiguous stimuli and short for unambiguous ones. However, the prototype concept, introduced to speech perception by Oden and Massaro (1978) and Repp (1976a) suggests that, even for stimuli that are consistently placed in the same category, there might be a gradient of reaction times reflecting their perceptual distance from the category prototype. The only attempt so far to test this hypothesis for stop consonants (Samuel, 1979) appears to have been unsuccessful. In other studies, too, labeling reaction times to different stop consonant stimuli strictly within the same category (if several such stimuli existed on a continuum) have tended to be equivalent (e.g., Pisoni & Tash, 1974).

Numerical ratings and reaction times have also been collected in discrimination tasks. Vinegrad (1972) conducted a direct magnitude scaling study with stop consonants (/bɛ/-/dɛ/-/gɛ/), vowels (/i/-/I/-/ɛ/), and pure tones varying in frequency. The stimuli were presented in AXB triads, and the subjects' task was to locate X in relation to A and B by marking a point on a line. A and B were always the extreme endpoint stimuli of the continuum, which made the procedure highly similar to that of Strange (1972), who presented only the middle stimuli. The results were very clear-cut: The stop consonants exhibited strongly categorical perception; different stimuli from within the same category were located in the same place. Vowels, on the other hand, gave more continuous results, as expected. The results for the tones were similar to those for the vowels; however, neither were perfectly continuous (see Section 5.3).

Category boundary effects for isolated vowels have also been obtained in studies where the subjects' task was to rate the perceived similarity of stimuli drawn from a continuum (e.g., Golusina, cited in Chistovich, 1971; Van Valin, 1976). Unless subjects are very carefully instructed to base their judgments on auditory stimulus properties alone, this task is likely to elicit a phonetic strategy.

Following an earlier study by Strange and Halwes (1971), Pisoni and Glanzman (1974) obtained confidence ratings for discrimination judgments of stop consonants (/ba/-/pa/) presented in AXB and 4IAX formats. There was a very straightforward monotonic relation between discrimination accuracy and confidence; in other words, subjects accurately postdicted their own success on each trial. While performance was not any better with confidence ratings than without, the correlation obtained does suggest, as Conway and Haggard (1971) had observed earlier, that subjects have at least statistical information about acoustic stimulus differences, in the form of subjective uncertainty. Seen in this way, the Pisoni and Glanzman results are equivalent to a previous demonstration by Studdert-Kennedy, Liberman, and Stevens (1964) that reaction times in a stop consonant ABX task were shortest for between-category comparisons, where discrimination was easiest, and longest for within-category comparisons. These observations also raise the possibility that, rather than directly accessing some auditory memory representations, subjects might base decisions about stimulus differences on estimates of their subjective uncertainty in phonetic categorization.

Most of the studies discussed in this section demanded an overt indication of subjects' awareness of intraphonemic stimulus differences. The results provided relatively little evidence of such awareness as far as stop consonants are concerned. On the other hand, there is overwhelming evidence that acoustic stimulus properties do have perceptual effects that listeners are not directly conscious of. Some of this evidence comes from same-different reaction time studies, which will be reviewed in Section 5.1, together with the role played by the perhaps most obvious factor influencing the detectability of acoustic differences--the physical size of the difference itself (i.e., the "step size" on a continuum). Other studies have shown that the magnitude of the selective adaptation effect depends on the precise acoustic properties of the adapting stimulus (e.g., McNabb, 1976a; Miller, 1977, 1981; Miller & Connine, 1980; Samuel, 1979) and that the perception of fused dichotic stimuli is sensitive to similar acoustic variables (e.g., Miller, 1977; Repp, 1976a, 1977). These and other studies show that the auditory properties of stop consonant stimuli play a significant role at early, precategorical stages of processing (as they must).

It remains for us to mention several studies that assessed listeners' sensitivity to within-category differences by monitoring some more immediate response of the organism than overt labeling. Studies of vocal imitation fall in this category because immediate repetition does not require categorization of a stimulus. Harris, Bastian, and Liberman (1961) showed long ago that imitation of stimuli from a /slIt/-/splIt/ continuum was strongly categorical; that is, subjects were unable to reproduce the precise closure durations of the stimuli and instead produced only two types of utterances. Of course, this result may reflect articulatory limitations or habits rather than (or as well as) an influence of categorical perception on the articulatory response. (The motor theory does not even distinguish these two possibilities, for categorical perception is hypothesized to derive from articulation.) For this reason, perhaps, initation has rarely been used in later studies of categorical perception. A phoneme boundary effect in the imitation of isolated vowels was reported by Chistovich, Fant, de Serpa-Leitão, and Tjernlund (1966), whereas imitations of vowel durations by American listeners (Bastian & Abramson, 1964) showed no effect of phonetic categorization (see also Section 5.2.5).

A more covert, physiologic response to auditory stimuli may be obtained from the surface of the skull in the form of evoked potentials. Dorman (1974) presented listeners with stop-consonant-vowel stimuli differing in VOT. At varying times during a train of stimuli, the standard stimulus (/ba/) changed to a different stimulus either within the same category or in a different category (/pa/). The N1-P2 component of the evoked potential (100-200 msec after stimulus onset) was significantly larger for between-category shifts than for within-category shifts, and the response to the latter did not differ from that to a no-change control. Dorman interpreted his results as reflecting immediate phonetic recoding.

Curiously, Dorman's results were not mentioned by Molfese (1978), who reinvestigated the problem using principal-components analysis of evoked-potential waveforms. His subjects listened to stimuli from a /ba/-/pa/ continuum and identified each stimulus by pressing one of two keys. The results were complex but suggested that within- as well as between-category differences affected the electric brain response. This basic finding was replicated with /ga/-/ka/ stimuli in 4-year-old children (Molfese & Hess, 1978) and 2- to 5-month-old infants (Molfese & Molfese, 1979). The evoked potentials of these young subjects also exhibited a component that responded only to between-category differences, while those of newborn infants did not (Molfese & Molfese, 1979), and those of adults (Molfese, 1978) followed a somewhat more complex pattern. These findings are intriguing, although they are not without methodological problems; at the simplest level of interpretation, they suggest that neuroelectric correlates of both auditory and phonetic processing may be found.

Changes in evoked potentials for within-category differences occur without the subject's awareness. However, some striking evidence that listeners _can_ gain conscious access to subphonemic acoustic stimulus differences comes from several studies that provided extensive training for the listeners. Although these results would fit in the present section on paradigms, we prefer to discuss them in Section 6, which deals with subject factors in categorical perception, one of which is experience.

5. STIMULUS FACTORS IN CATEGORICAL PERCEPTION

In this section, we will review various relevant factors residing in the stimuli themselves (rather than in their arrangement or in the kinds of responses given by subjects). In Section 5.1, we will examine the effects of variables operating within a given set of stimuli, the most important ones being physical separation (step size) and duration. In Section 5.2, we will review differences in the degree of categorical perception among different stimulus sets, focusing on stimuli other than the ubiquitous stop consonants and vowels. This will lead us to a detailed consideration of the perception of "nonspeech analogs" of speech stimuli, together with findings of categorical perception of other kinds of nonspeech stimuli (Section 5.3).

5.1. Stimulus Factors and Auditory Memory

5.1.1. Step Size Effects

The variable most obviously related to the ease of discriminating two stimuli is the magnitude of the physical difference. Several levels of this variable, in the form of different "step sizes" in comparisons drawn from a continuum, have been included in most studies of categorical perception, including the earliest ones. It is a commonplace finding that 2-step discrimination performance is higher than 1-step discrimination performance, 3-step is higher than 2-step, and so on. One might think that here is _prima facie_ evidence that listeners are sensitive to subphonemic physical differences between the stimuli. However, the issue is not that simple: Stimuli that are more widely separated on the physical continuum generally are more likely to be classified into different categories, and under the assumption that discrimination is mediated by category labels, discrimination accuracy is _predicted_ to increase with step size. Therefore, an effect of step size cannot be taken to reflect auditory (rather than phonetic) discrimination unless it is significantly larger than predicted from (in-context!) labeling probabilities.

This point was given systematic attention by Healy and Repp (1982), who computed the differences between predicted (in-context) and obtained "same-different" discrimination performance at three different step sizes for four different stimulus continua (stop-consonant-vowel syllables, isolated vowels, isolated fricative noises, and complex tones varying in timbre). The idea was that, given a linear measure of performance (d' in their case; percentages are not suitable because of their inherent nonlinearity), the predicted-obtained differences should increase with step size if listeners are indeed sensitive to acoustic differences; otherwise, the step size effect should be fully accounted for by the in-context predictions from labeling performance. Healy and Repp found that a residual step size effect was present for vowels and tones, and probably for fricative noises as well (a ceiling effect prevented statistical significance), but not for stop consonants. Since stop consonant discrimination was generally slightly worse than predicted (a seemingly unusual result that, however, reflected the effective partialling out of contrast effects in labeling), the results provided strong support for the hypothesis that stop consonant discrimination was based exclusively on phonetic labels. Apparently, the subjects in the Healy-Repp experiment retained no distinctive acoustic details of stop consonant stimuli but did make use of auditory information with the other stimulus classes.

However, these results do not warrant the conclusion that acoustic properties of stop consonants do not enter auditory memory at all. Rather, their auditory traces may be so weak as to influence performance only under very special conditions. One sufficiently sensitive measure of performance appears to be reaction time in a same-different task. Pisoni and Tash (1974) adapted to speech perception a procedure used by Posner (e.g., Posner & Mitchell, 1967) in his well-known letter matching studies: A "same" judgment for two physically identical stimuli ("physical match") might be faster than a "same" judgment for two physically different stimuli from the same category ("name match"), if any auditory information is retained from the first stimulus in the pair. Similarly, "different" reaction times to two stimuli from opposite sides of a category boundary might be faster when the physical separation between the two stimuli is large than when it is small. Both results were reported by Pisoni and Tash (1974) for syllables from a /ba/-/pa/ continuum presented in pairs with 250-msec ISIs: When two stimuli from the same category were separated by two steps on the continuum, "same" responses were significantly slower than for pairs of identical stimuli; at the same time, subjects were not any more likely to say "different" to two-step pairs than to identical pairs, so that, overtly, perception was highly categorical. "Different" response latencies to stimuli crossing the boundary and separated by two steps were longer than for stimuli separated by four or six steps. However, there was no significant difference between four- and six-step "different" pairs and, moreover, the likelihood of incorrect "same" responses was highest for two-step pairs, so that the "different" reaction times may have reflected uncertainty in phonetic, rather than auditory, judgments.

On the basis of their results, Pisoni and Tash (1974) proposed a two-stage model for same-different comparisons, according to which a comparison of auditory stimulus properties precedes the comparison of phonetic labels, the second stage being used only if the auditory difference falls neither below the "same" nor above the "different" criterion adopted by the listener. This ordering of stages is reversed with respect to the Fujisaki-Kawashima dual-process model for ABX discrimination, which puts the phonetic comparison first. However, unlike the Pisoni-Tash model, the Fujisaki-Kawashima model was not intended to describe real-time information processing; rather, it merely captures the fact that phonetic categories loom large in the listener's awareness and actually permits either order of deployment of the two component processes.

The demonstration by Pisoni and Tash that some acoustic properties of stop consonants are retained in memory inspired other researchers to ask whether these memory traces, like those of isolated vowels, decay over time.

Several studies addressing this question have yielded mixed results. Eimas and Miller (1975) presented pairs of stimuli from a /ba/-/da/ (formant transition) continuum at three ISIs (50, 200, and 800 msec). Since the distinctive information was located at stimulus onset, stimulus onset asynchrony (SOA) is a more appropriate measure of temporal separation; the SOAs were 310, 460, and 1060 msec. "Same" latencies were significantly faster for physically identical stimulus pairs than for physically different pairs, but only at the two shorter SOAs. At the shortest SOA (310 msec), subjects actually detected the physical within-category difference on 22.8 percent of the trials, as compared to 2.8 percent at the 460-msec SOA. A partial replication of these results was obtained in a second study by Eimas and Miller (1975) with a /ra/-/la/ continuum. These findings provided rather striking support for a rapidly decaying auditory memory that, after 460 msec, no longer afforded conscious detection of within-category differences but still generated a reaction time difference that disappeared after 1060 msec. The fast decay of the memory relative to the 3-sec asymptote found in studies with vowels (see Section 4.1.2) may reflect the initial "weakness" of the auditory trace (i.e., the general auditory similarity of the stimuli in the set--cf. Darwin & Baddeley, 1974). It should be added that the data of Eimas and Miller, like those of Pisoni and Tash, did not yield any unambiguous evidence for any involvement of auditory memory in "different" judgments.

Negative results were obtained in two unpublished studies by Repp (1975, 1976b). Repp (1975) used /ba/-/pa/ stimuli similar to those of Pisoni and Tash (1974) and presented them to different ears at a number of SOAs ranging from 0 to 3.3 sec. The listeners were given two types of instruction: Either they were told to make their same-different judgments on the basis of stimulus categories only (phonetic matching condition), or they were given some experience with the stimulus continuum (following the example of Pisoni & Lazarus, 1974) and then tried to make auditory same-different judgments (physical matching condition). The expected effect of physical mismatch on "same" latencies was only weakly present in the phonetic matching condition and did *not* systematically decline with SOA; it was totally absent in the auditory matching condition where subjects, surprisingly, proved *less* sensitive to physical differences than in the phonetic matching condition. Thus, this study provided no evidence whatsoever for auditory memory. Perhaps, presentation of the stimuli to different ears prevented the efficient use of auditory memory. In an attempt to examine this possibility, Repp (1976b) presented stimuli either binaurally or to different ears at one of two SOAS, 500 or 2000 msec. By using only four different stimuli (/bæ/, two versions of /dæ/, and /gæ/), Repp controlled for the effect of labeling uncertainty on reaction times, thereby making "different" latencies a potentially unconfounded indicator of auditory memory. However, the results of this study were entirely negative: There were no significant step size effects in either "same" or "different" latencies.

Another study in the same vein, and the only one to be published, was conducted by Hanson (1977). Like Repp (1975), she used a /ba/-/pa/ continuum and two different sets of instructions (phonetic matching and physical matching). Unlike Repp, she presented her stimuli binaurally and had only two SOAs, 550 and 870 msec, which were varied between subjects. Although Hanson was successful in eliciting better discrimination performance through physical matching instructions (see Section 6.1.2), step size effects were absent in the physical matching task and only weakly present in the phonetic matching task. Hanson's study must be viewed with caution because of high error rates and because it is the only study in the literature that failed to find a reaction time peak at the category boundary in a simple labeling task.

In summary, same-different reaction time studies have yielded some rather clear instances of listener sensitivity to within-category differences among stop consonant stimuli, but there are also failures to obtain such effects. While the causes of the negative findings remain obscure, the positive results do strengthen the hypothesis that all aspects of speech signals are represented in auditory memory.

5.1.2. Stimulus Duration

We turn now to a group of studies that attempted to either increase or decrease categorical perception by directly manipulating the stimuli, with the purpose of thereby modifying the strength of their auditory memory representations. One manipulation that promised to have some effect was to vary stimulus duration. In the case of homogeneous stimuli, such as the steady-state vowels used in a number of experiments, a reduction in stimulus duration might weaken the auditory trace and thereby lead to more nearly categorical perception.

The first study to test this hypothesis was conducted by Fujisaki and Kawashima (1968). They presented vowels from an /i/-/e/ continuum (there is no /I/ category in Japanese) in identification and ABX discrimination tasks, with stimulus duration set at either 25, 50, or 100 msec. A subsequent paper (Fujisaki & Kawashima, 1969) reports data from a similar experiment with shorter vowel durations--1, 3, or 6 pitch pulses, corresponding to durations of 8, 23, and 46 msec. Finally, Fujisaki and Kawashima (1970) presented what seem to be new data for single-pulse (8 msec) and 100-msec vowels. In all three reports, the figures show that discrimination performance was (paradoxically) higher for the short vowels, while the accompanying text consistently states the opposite. These inconsistencies in the Fujisaki-Kawashima papers were apparently not noticed by other authors concerned with the same issue: Pisoni (1971, 1973, 1975) paid attention only to the text, while Tartter (1982) paid attention only to the figures. In the light of Pisoni's later findings, the only plausible explanation is that Fujisaki and Kawashima kept using incorrect figure legends, and that their data really showed what they claimed to have found--namely, poorer discrimination and more nearly categorical perception of short vowel stimuli.

Pisoni (1971) investigated the matter more systematically. In his Experiment III, he presented short (50 msec) and long (300 msec) vowels from an /i/-/I/ continuum in identification and ABX discrimination tasks. Although this preliminary study involved only five subjects, it did yield significantly (but not dramatically) higher discrimination scores for the long vowels. A replication with a larger number of subjects was reported by Pisoni (1975, Exp. I). Again, performance was slightly higher for the long vowels, but the difference reached significance only for 1-step, not for 2-step comparisons.

In another experiment, Pisoni (1971, Exp. IV) presented short (50 msec) and long (300 msec) vowels from an /i/-/I/-/ɛ/ continuum in identification, ABX, and 4IAX tasks. Besides getting substantially higher and virtually continuous discrimination performance in the 4IAX paradigm, he also obtained consistent differences in favor of the long vowels, which were especially clear in the 4IAX test. A replication using an /i/-/I/ continuum was conducted by Pisoni (1975, Exp. II), which again yielded sizeable effects of vowel duration (although they were, surprisingly, reported to be statistically nonsignificant).

Vowels of different duration were also used in Pisoni's (1971: Exp. VI, 1973) study of same-different discrimination at different temporal delays, and while there was little difference on "between-category" trials, performance for long vowels was clearly higher on "within-category trials," where auditory memory was presumed to be the prime source of distinctive information. Similar results were obtained by Sachs (1969), who used 150-msec and 250-msec /a/-/æ/ vowels in an absolute identification task. Tartter (1982), in a recent critical review, overlooked these data when she concluded that changes in vowel duration have equal effects across a vowel continuum and that, therefore, the dual-process model should be rejected. While the data reviewed in the preceding two paragraphs indeed showed fairly uniform effects of vowel duration across a continuum, those just cited do support the dual-process model by showing that perception of short vowels is more nearly categorical (especially at long interstimulus intervals) than perception of long vowels. Because the gradual transitions between categories make it difficult to achieve a clear separation of between- and within-category pairs on a vowel continuum, the inconsistencies in the literature with regard to the uniformity or nonuniformity of performance decrements across a continuum can hardly justify the rejection of a model as conceptually sound as the dual-process model. It is possible, however, that the influence of phonetic categorization on vowel doscrimination is more indirect than is generally assumed (see Section 4.1.2).

Vowel duration effects have also been obtained in verbal memory research: Crowder (1973a) found that the suffix effect was smaller for lists of short vowels than for lists of long vowels. It has also been reported that shortened vowels exhibit a right-ear advantage in dichotic presentation while long vowels do not (Godfrey, 1974). All these results strongly suggest that auditory memory strength depends on the duration of a (homogeneous) stimulus.

A more radical modification of vowel duration was recently performed by Tartter (1981). She started with stimuli from an /I/-/ɛ/ continuum, 260 msec in duration, and obtained typical identification and oddity discrimination functions. Then she preceded the stimuli with 40-msec formant transitions appropriate for /b/. In one condition, the transitions for each vowel started at the same frequencies; in a second condition, they started at different frequencies that covaried with the vowel steady-state frequencies, so that transition slopes remained constant. Neither manipulation had any effect on vowel discrimination--not an unexpected finding in view of the poor auditory memory for transitional cues on stop consonant continua (e.g., Pisoni, 1971). In a subsequent condition, however, Tartter removed the vocalic steady states, leaving only the 40-msec transitional portions. The vowels were still identified quite accurately from these truncated /b/-vowel syllables, but discrimination performance suffered considerably. For both sets of transitions, perception was virtually categorical, and the results exhibited the pattern typical for stop consonant continua. This finding strongly suggests that rapidly changing acoustic information is poorly retained in auditory memory, regardless of whether it conveys consonantal or vocalic distinctions, and that the noncategorical perception of isolated vowels is due to their steady-state characteristics and their resulting salience in auditory memory, not to any special perceptual status of vowels as phonological segments.

This conclusion is further supported by the results of studies on the perception of vowels in context (Sachs, 1969; Stevens, 1968). The stimuli in these studies were not simply steady-state vowels embedded in some acoustic context (as they are sometimes described in the literature) but synthetic words with little (Sachs) or no (Stevens) steady-state vocalic portion. In Stevens' (1968) study, the continuum ranged from /bil/ (a nonsense word) to /bIl/ and was obtained by interpolating between formant patterns obtained from natural utterances. Listeners actually perceived three categories ("beel," "bill," and "bell") but, in an ABX test, showed sharp discrimination peaks at both category boundaries, indicating strongly categorical perception. A matched continuum of isolated steady-state vowels was included as control and yielded results typical of noncategorial perception.

Sachs (1969) employed a /bɑdal/-/bædəl/ (or "bottle"-"battle") continuum together with two matched steady-state /ɑ/-/æ/ continua of different durations. Measuring discrimination by computing d' indices for pairs of adjacent stimuli from the results of an absolute identification task, he found a pronounced peak at the category boundary for the word continuum, a somewhat less pronounced peak for the short vowels, and even less of a peak for the long vowels. Although neither Stevens nor Sachs compared their discrimination data to predictions generated by the Haskins model, the pattern of their results suggests fairly categorical perception of vowels in word context. A recent study by Sawusch, Nusbaum, and Schwab (1980) yielded similar results. They used /i/-/I/, /sis/-/sIs/, and /bit/-/bIt/ continua and obtained more nearly (though not completely) categorical results for the latter two. The fact that they observed no difference between the two context conditions, one of which merely put steady-state vowels in a fixed fricative-noise context while the other contained time-varying vocalic portions, suggests that auditory memory may be weakened by either dynamic change or by the presence of irrelevant context.

The finding of increased categorical perception for shortened or dynamically varying vowels suggests that the short duration and rapidly changing nature of the critical cues for initial stop consonants may be at least partially responsible for their categorical perception. One way to investigate this hypothesis with stop consonant stimuli is to lengthen (and, thereby, also to slow down) the formant transitions that distinguish different places of articulation. This was done in two nearly simultaneous but independent studies by Dechovitz and Mandler (1977) and by Keating and Blumstein (1978). Dechovitz and Mandler extended the F2 and F3 transitions of a /ba/-/da/-/ga/ continuum from 30 to 135 msec. It was known from informal observations that a syllable with such extended transitions sounds rather similar to the original, as long as the F1 transition remains constant. This impression was confirmed by the results of identification and same-different discrimination tests that showed no difference between the original and extended-transition stimuli: Perception of both sets of stimuli was strikingly categorical.

Keating and Blumstein (1978) used a /da/-/ga/ continuum with three lengths of F2 and F3 transitions (45, 95, and 145 msec). The three sets of stimuli yielded similar results in identification and 4IAX discrimination tests, although there were some significant differences, primarily due to the stimuli with intermediate transition length, which were discriminated best. Within-category discrimination in this study was significantly better than predicted (perhaps due to the sensitive 4IAX paradigm), particularly with the longer transitions. Therefore, the Keating and Blumstein results are not entirely negative, but they do suggest that the short duration of F2 or F3 transitions is not a major determinant of categorical perception.

A very interesting result was recently reported by Tartter (1981). She removed the steady-state vocalic portions of /ba/-/da/ stimuli, leaving only the initial 40 msec that contained the formant transitions. Compared to the full syllables, this resulted in a distinct improvement in within-category discrimination (an oddity task was used), while stop consonant identification was just as accurate as when the steady states were present. This finding strongly suggests that the formant transitions have a representation in auditory memory that can be accessed when the redundant steady state is eliminated. Thus, the vocalic portion of a stop-consonant-vowel syllable, while it aids phonetic perception, appears to interfere with the preservation of consonantal cues at a precategorical level. The overriding auditory salience of an irrelevant stimulus portion may be a major factor causing categorical perception.

5.1.3. Other Stimulus Parameters That May Affect Categorical Perception

One parameter that generally has received little attention in speech perception research is amplitude. However, recent studies by Syrdal-Lasky (1978), Dorman and Dougherty (1981), and Van Tasell and Crump (1981) have shown that the identification of synthetic stop consonants varying along a place-of-articulation continuum may exhibit large shifts with changes in playback level. Syrdal-Lasky also presented her stimuli in an oddity discrimination task and found different discrimination functions at different signal levels. However, it seems from an inspection of her figures that, if the changes in labeling probabilities are taken into account, perception was about equally categorical in all conditions. It is tempting to speculate that auditory discrimination along some physical dimension might be improved when that dimension is highlighted by increasing its amplitude relative to nondistinctive signal components. However, so far there are no data pertaining to this hypothesis.

Another parameter that does not seem to have much effect on categorical perception is whether a stimulus is periodic or aperiodic, other things equal. Fujisaki and Kawashima (1968) synthesized an /i/-/e/ continuum with either periodic or aperiodic excitation. There was a shift in the category boundary (more /i/ responses were given to the aperiodic vowels) and ABX discrimination functions showed a corresponding peak shift but did not differ in overall level. Highly similar (though not completely identical) data were reported by Fujisaki and Kawashima (1969). Thus, periodicity, like overall amplitude, seems to affect categorical perception only to the extent that labeling probabilities are affected; these variables do not seem to have any direct influence on the strength of the auditory trace. This conclusion was further supported by a recent study by May and Repp (1982), who failed to find any difference in auditory memory for periodic and aperiodic nonspeech stimuli (single-formant resonances).

One stimulus factor that has not been systematically investigated but may well play a role in categorical perception is naturalness. Poorly synthesized stimuli may be expected to be less categorically perceived (given that they are sufficiently distinct acoustically) than good synthetic stimuli or natural speech. The reason for this is that poor stimuli may make it easier for listeners to adopt auditory strategies in discrimination, while highly realistic stimuli may elicit a phonetic strategy. (More about strategies in Section 6.1.)

5.2. Different Classes of Speech Sounds

The large majority of studies concerned with categorical perception and related topics have used as materials either the two standard sets of prevocalic stop consonants (VOT or place-of-articulation continua) or isolated steady-state vowels. In this subsection, we will review studies that examined other types of speech contrasts or used less common varieties of stop consonant or vowel continua. We will pay some attention to the specific stimulus parameters that were varied to obtain a continuum, as these may have a bearing on the strength of the auditory memory trace.

5.2.1. Stop Consonants

Voicing continua. The earliest voicing continua were generated on the Haskins Laboratories Pattern Playback by the procedure called "F1 cutback"--increasing delays in the onset of F1 relative to the onsets of the higher formants. Perception of these stimuli was highly categorical (Liberman, Harris, Kinney, & Lane, 1961). During the following years, Abramson and Lisker developed the now commonly used procedure for varying VOT, which combines a delay in the onset of F1 with the substitution of aperiodic for periodic energy in the higher formants during the period of the delay. These stimuli, too, show highly categorical perception in the standard experimental setup (Abramson & Lisker, 1970; Lisker & Abramson, 1970). The original Abramson-Lisker stimuli, which have been used in many different studies, included variations in VOT on the "negative" side: Different degrees of prevoicing were simulated by preceding the stop release with varying amounts of low-energy buzz from the periodic source of the synthesizer. This region of the continuum is of interest because prevoicing is not distinctive in English (and native speakers of English are very poor in discriminating differences in prevoicing--cf. Abramson & Lisker, 1970), while it is in some other languages (see Section 6.2).

In acoustic terms, the Abramson-Lisker VOT continuum is really not one continuum but two: The acoustic variations used to achieve different degrees of prevoicing (voicing lead) are quite different from those used to generate different degrees of aspiration (voicing lag). On the "positive" side, as increasing amounts of aspiration are substituted for voicing, there is at first a correlated spectral change as the F1 transition (always rising) is cut back more and more, so that the onset of F1 occurs at increasingly higher frequencies and amplitudes. Spectral cues, particularly from the F1 region, are relevant to the perception of voicing, as several studies have shown (Lisker, Liberman, Erickson, Dechovitz, & Mandler, 1977; Stevens & Klatt, 1974; Summerfield & Haggard, 1977). As voicing onset is delayed beyond the region of the formant transitions (the first 30-70 msec), the spectral covariation ceases but the duration of the periodic portion decreases as the aspirated position increases. This negative covariation has been given little attention in the past, although it may play a role when VOTs get rather long and the periodic portions short enough for the temporal variations to exceed the detection threshold (cf. Wood, 1976a). An alternative, and perhaps preferable, way of synthesizing VOT continua in the long positive range would be to hold the duration of the periodic portion constant (cf. Repp, 1981b).

A procedure for generating VOT continua (in the positive VOT range) by cross-splicing pitch periods and aspiration from natural-speech tokens was devised by Lisker (1976) and described in detail by Ganong (1980). There is little doubt that such stimuli are perceived categorically: Repp (1981b, Exp. 3) presented stimuli from a natural-speech VOT continuum in a fixed-standard AX task and obtained extremely poor within-category discrimination performance.

The highly categorical perception of stop consonant voicing in initial position may be contrasted with the less categorical perception of the same phonetic distinction in final position. This comparison is important, as it shows that categorical perception is not only a function of phonological status but also of the acoustic stimulus dimensions varied. One important cue for consonant voicing in postvocalic position (in English) is the duration of the vocalic portion. Using variations in "vowel duration" to generate a variety of voiceless-voiced continua (including final fricatives and stop-fricative clusters as well as final stops), Raphael (1972) found that oddity discrimination was much better than predicted, given a sufficiently large physical difference. There also appeared to be a discrimination peak at the category boundary, making the data similar to those typically obtained with isolated vowels. Although there have been numerous studies of the various cues to the voicing distinction in postvocalic position, Raphael's remains the only study to date that included discrimination tests.

The voicing contrast for stops in intervocalic position may be cued by variations in the duration of the (silent) closure interval. Liberman, Harris, Eimas, Lisker, and Bastian (1961) synthesized a /ræbId/-/ræpId/ continuum in this way and presented it in identification and ABX discrimination tasks. The results provided an interesting instance of perception that was neither very categorical nor very continuous: Discrimination performance was considerably better than predicted but showed a peak at the boundary. A second peak was noted within the "p" category and attributed to subjects' use of a covert third category, "unnatural 'p'." However, even revised predictions based on three categories did not reach the level of the obtained discrimination performance. Here is a case, it seems, where the contribution of phonetic and auditory processes to discrimination were in approximate balance.

Place-of-articulation continua. Early studies used two-formant stimuli in which the F2 transition was the sole cue to place of articulation (Liberman et al., 1957; Mattingly et al., 1971). Despite the relative crudeness of the stimuli, the perception of these syllable-initial stops was invariably quite categorical. Later experiments in which stimuli also had a varying F3 transition yielded similar results (e.g., Pisoni, 1971). Numerous studies have employed variants of /b/-/d/-/g/ continua, and the categorical discrimination of these stimuli is one of the most consistently replicated results in speech perception research, notwithstanding Barclay's (1972) findings (see Section 4.2). All of these studies used formant transitions as the sole cue to place of articulation; so far, the discriminability of variations in release burst spectrum (another important cue for stop consonant place of articulation) has not been tested. Also, there are very few studies that have employed continua of voiceless stops (/p/-/t/-/k/). What data there are (Syrdal-Lasky, 1978, used F1 cutback without aspiration) suggest categorical perception.

Syllable-final stops varying in place of articulation were synthesized by Mattingly et al. (1971) by varying the final F2 transition in two-formant stimuli (/ab/-/ad/-/ag/). The oddity discrimination function for these sounds showed no clear peaks at phonetic boundaries, which the authors attributed to the poor quality of the stimuli. Subsequently, Popper (1972) found a well-defined peak on an /ab/-/ad/ continuum, but within-category same-different discrimination was better than predicted by the Haskins model. Recently, Miller, Eimas, and Zatorre (1979) obtained similar results with /ab/-/ad/ stimuli in an oddity discrimination task: There was a discrimination peak at the category boundary but also unexpectedly high performance within the /ad/ category, which the authors were unable to explain. Taken together, these results suggest that syllable-final stops are not perceived as categorically as syllable-initial stops. One likely reason is that the distinctive information, being in final position, is better retained in auditory memory. (Cf. the importance of offset frequency in determining the pitch of nonspeech frequency glides--e.g., Brady, House, & Stevens, 1961; Schwab, 1981.) However, one study that directly compared initial and final stops (Larkey et al., 1978), using stimuli that were acoustic mirror images, found equally categorical perception for both.

Manner continua. One primary cue for the perceived presence or absence of a stop consonant in medial position is the presence or absence of an appropriate closure interval. Bastian et al. (1961) constructed a continuum from /slIt/ to /splIt/ by inserting increasing amounts of silence after the /s/ noise of a natural-speech token of /slIt/. The stimuli were presented in identification and oddity discrimination tasks, and the listeners' responses proved to be highly categorical, with obtained discrimination scores only slightly exceeding the predictions of the Haskins model. These results were essentially replicated in a recent study by Fitch et al. (1980) with a synthetic /slIt/-/split/ continuum, although these authors did not conduct a direct comparison of predicted and obtained discrimination scores. Even more recently, Best et al. (1981) presented a synthetic /sei/-/stei/ continuum, generated similarly by varying silent closure duration, in oddity and same-different tasks and also computed the Haskins model predictions. The discrimination functions showed pronounced peaks at the category boundary, but performance in both tasks was a good deal better than predicted, particularly within categories. Thus, in this study the listeners did seem to pick up some auditory differences. Also, Repp (1981b) recently obtained rather good within-category discrimination of closure duration differences in /splIt/ and /stei/ stimuli in a fixed-standard AX task.

A related stop manner contrast is that between a fricative and an affricate (effectively, stop + fricative). In intervocalic position, this difference may be cued by silence preceding the fricative noise (e.g., Gerstman, 1957). Employing stimuli from a "say shop"-"say chop" continuum in a fixed-standard AX discrimination task, Repp (1981b) obtained fairly high within-category discrimination, which adds to the mounting evidence that within-category differences in temporal stimulus structure are detected more readily than differences in spectral structure. Another way of cueing the fricative-affricate distinction is by means of fricative noise duration (Gerstman, 1957), but no discrimination data for this cue are in the literature. A third important cue is the amplitude rise time of the noise, and this cue has been investigated in initial position by Cutting and Rosner (1974, 1976). They generated synthetic /tʃa/-/ʃɑ/ and /tʃæ/-/ʃæ/ continua by varying the rise time of the fricative noise, and presented the stimuli in identification and ABX discrimination tasks. The results showed fairly categorical perception, even though fricative noise duration apparently covaried with rise time.

5.2.2. Nasal Consonants

Nasal consonants are relative late-comers on the scene because it took some time before convincing nasals could be produced synthetically. Initial studies by Garcia (1966, 1967a, 1967b) still suffered from stimulus problems. She (Garcia, 1966) converted a two-formant /bɛ/-/dɛ/-/gɛ/ continuum into a /mɛ/-/nɛ/-/ŋɛ/ continuum by simply preceding the stimuli by a constant synthetic nasal murmur. An /ɛm/-/ɛn/-/ɛŋ/ continuum was obtained by playing the stimuli backwards. It turned out that the nasals were labeled rather poorly, especially in initial position. Discrimination performance was also rather poor, but did show some evidence of peaks at category boundaries for subjects who labeled the final nasals consistently. Somewhat more consistent data were obtained in a replication with three-formant stimuli (Garcia, 1967a, 1967b). They suggested fairly categorical perception.

Much cleaner results were obtained by Miller and Eimas (1977), who compared a /ba/-/da/ with a /ma/-/na/ continuum, obtained by adding initial nasal resonances and by flattening the F1 transition. Although the nasal categories were not quite as sharply separated as the stop categories, discrimination of both stimulus sets was equally categorical in an oddity task, with obtained scores only slightly better than predicted. A careful replication of Garcia's work was undertaken by Larkey et al. (1978), who not only used all three nasal categories in initial and final position (with the vowel /æ/), but also compared their perception with that of matched stop consonant continua. The results showed highly categorical perception of all stimulus sets, with somewhat better within-category discrimination for final than initial nasals. In the meantime, Miller and Eimas also extended their study to syllable-final nasals (Miller et al., 1979) and obtained categorical perception, except for high levels of discrimination within the /n/ category. In view of the Larkey et al. data, this is likely to have been a stimulus artifact of some sort.

Given the consistently categorical results for both stop consonants and nasals, the results of experiments using stop-to-nasal (oral-nasal) continua would seem highly predictable. Yet, these studies are not trivial, for the acoustic dimension cueing the oral-nasal distinction (amplitude or duration of nasal resonance) is considerably less complex and, therefore, perhaps more readily discriminable than the spectral changes cueing place-of-articulation distinctions. Thus, oral-nasal continua offer an opportunity for noncategorical perception, even though the phonetic boundary may coincide with the auditory detection threshold for the presence of nasal murmur. The first study was conducted by Mandler (1976), who synthesized /ba/-/ma/ and /da/-/na/ continua by two different methods, using either the oral branch or the nasal branch of a serial resonance synthesizer. In each case, the amplitude of the simulated nasal resonance was varied in a number of steps. The labeling functions for these continua were not very steep, but same-different discrimination scores showed a peak in the boundary region, suggesting categorical perception.

Rather similar results were obtained by Miller and Eimas (1977) for synthetic /ba/-/ma/ and /da/-/na/ continua obtained by simultaneously varying the duration of nasal murmur and F1 onset frequency (which is higher for nasal than for oral stops). Again, labeling functions were rather gradual, but oddity discrimination functions exhibited peaks. Discrimination was somewhat better than predicted. (An unusually high level of discrimination performance in comparisons involving the most stop-like stimulus was traced to a stimulus artifact and eliminated in a supplementary experiment, described in the same paper.) Equally categorical perception was found for syllable-final /ab/-/am/ and /an/-/ad/ continua (acoustic mirror-images of the original stimuli) by Miller et al. (1979).

A possibility suggested by the motor theory of speech perception is that categorical-like perception might be caused by a nonlinear relation of an acoustic continuum to changes along the corresponding articulatory dimension. In the case of the oral-nasal distinction, this problem was addressed by Abramson, Nye, Henderson, and Marshall (1981), who created a /da/-/na/ continuum on an articulatory synthesizer by directly controlling the degree of velar opening. The amplitude of nasal murmur was determined to be a negatively accelerated function of the velopharyngeal port area, which was varied in equal steps. While the category boundary was once again not very sharp, AXB discrimination functions showed clear peaks that unmistakably pointed towards categorical perception, even though no predictions were calculated. Thus, the observed nonlinear relation between articulation and acoustic output was not responsible for categorical perception in this instance.

5.2.3. Liquids and Semivowels

In a study primarily intended to demonstrate effects of linguistic experience (see Section 6.2), Miyawaki et al. (1975) synthesized a /ra/-/la/ continuum by varying the onset frequency of F3, which, in this instance, had an initial 50-msec steady state followed by a 75-msec transition. American listeners perceived the stimuli fairly categorically: Oddity discrimination scores showed a clear peak at the boundary, but within-category discrimination was significantly better than predicted, particularly within the /la/ category. Clearly, perception was less categorical than that of stop consonants. McGovern and Strange (1977) subsequently conducted experiments with synthetic, mirror-image /ri/-/li/ and /ir/-/il/ continua and obtained results very similar to those of Miyawaki et al. So did MacKain et al. (in press) with a /rak/-/lak/ continuum in AXB and oddity discrimination tests.

Fujisaki and Kawashima (1970) obtained a (Japanese) /wa/-/ra/ continuum by varying the frequency of the (rather slow) F2 transition. ABX discrimination functions showed a broad peak at the category boundary--considerably broader than predicted. Thus, perception of this continuum was not highly categorical. More nearly categorical results were obtained by Frazier (1976), who synthesized an acoustic continuum from /wɛ/ to /lɛ/ to /yɛ/ by varying the initial steady state (90 msec) and transition (60 msec) of F2. A mirror-image /ɛw/-/ɛl/-/ɛy/ continuum was also used. The stimuli were presented in identification and same-different discrimination tests at two different ISIs (57 msec and 1 sec). The results revealed highly categorical perception in all conditions. The ISI seemed to have no effect on performance.

Miller (1980) has reported essentially categorical perception of stimuli from a stop-semivowel continuum (/ba/-/wa/), obtained by varying the duration of the initial formant transitions (Miller & Liberman, 1979). This study also demonstrated a shift in the discrimination peak along with a shift in the category boundary when the duration of the steady-state vocalic portion was extended. (However, this shift may have a purely psychoacoustic reason--see Carrell, Pisoni, & Gans, 1980.) More recently, Godfrey and Millay (1981) found somewhat less categorical perception of a /bɛ/-/wɛ/ continuum, due to rather high discrimination scores within the /b/ category.

5.2.4. Fricatives

Fricative consonants offer a better opportunity for noncategorical perception than any speech sounds discussed so far in this section. Fricative-vowel stimuli contain a noise portion that is nearly homogeneous, lasts for 100 msec or more, and has a characteristic "pitch." Moreover, stimuli along a synthetic fricative continuum tend to be rather widely spaced, so that even 1-step differences should exceed the auditory detection threshold.

The first categorical perception study with fricatives was conducted by Fujisaki and Kawashima (1968). They synthesized a /ʃ/-/s/ continuum by varying the frequencies of two fricative poles (formants) and presented these noises either in isolation or followed by a vowel (probably /e/--cf. Fujisaki & Kawashima, 1970). The ABX discrimination results were rather variable and showed fairly good within-category discrimination, especially at the /ʃ/ end, but there was also a peak at the category boundary. The vocalic context depressed discrimination scores somewhat, without changing the shape of the discrimination function. Fujisaki and Kawashima (1969) report slightly different data for the same experiment. (Perhaps, subjects had been added.) However, there was no consistent effect of vowel context. Finally, Fujisaki and Kawashima (1970) display yet another set of data, again showing peaks at the boundary, but now better within-category discrimination in vocalic context. Thus, while the effect of context is not clear at all, the data consistently show moderately categorical perception of fricative noises in context and in isolation. The finding for isolated noises contrasts starkly with results obtained by Healy and Repp (1982), who found discrimination in a same-different task to be essentially continuous. However, Healy and Repp used larger step sizes than Fujisaki and Kawashima, and a ceiling effect may have obscured a possible discrimination peak at the boundary. The high scores achieved by subjects at larger step sizes show quite clearly, however, that acoustic differences between isolated fricative noises are not hard to detect (cf. also Repp, 1981c). The perception of these stimuli appears to be at least as noncategorical as that of isolated vowels.

Fricatives in vocalic context also have yielded conflicting results. A dissertation by Hasegawa (1976) examined noises from a /ʃ/-/s/ continuum in postvocalic position, following either /i/ or /u/. The subjects were first given considerable training in ABX discrimination of vowels. Their fricative discrimination was essentially continuous; there was not even a hint of a peak at the category boundary. May (1981), on the other hand, obtained fairly categorical perception for three fricative continua presented to Egyptian listeners in a 4IAX paradigm. The continua ranged from /ʃ/ to /s/, from /x/ to /ħ/, and from /ɣ/ to /ʕ/, always in intervocalic context (/ə-ə/). While discrimination performance was better than predicted, all three continua showed a discrimination peak at the boundary. Repp (1981c) recently synthesized /ʃa/-/sa/ and /ʃu/-/su/ continua and presented them in AXB and fixed-standard AX tasks. In both tasks, the majority of subjects perceived the stimuli quite categorically: Although within-category discrimination was better than predicted, the peaks at the category boundary were extremely pronounced. However, there were some subjects whose discrimination scores were far superior and probably continuous. (A ceiling effect prevented any peaks from appearing.) These subjects apparently followed a radically different perceptual strategy. (See Section 6.1 for further discussion.) Fricative stimuli seem to be especially suited for the application of different strategies, so that they may be perceived fairly categorically in one situation but continuously in another. This may explain the conflicting results in the literature.

5.2.5. Vowels

Most of the vowel studies in the literature have already been reviewed in Section 4 or will be reviewed in Section 6. We note here that the finding of a discrimination peak at the category boundary is the rule rather than the exception; the earliest study by Fry et al. (1962) is one of the few that did not find a peak. We also note that most studies used continua of high front vowels (the /i/-/e/ range). The instability of vowel category boundaries and the magnitude of context effects in labeling may be due in part to the inclusion of categories such as /I/, which do not normally apply to isolated vowels (cf. Strange, Edman, & Jenkins, 1979). While the primary reason for the noncategorical perception of isolated vowels is undoubtedly their inherent high discriminability and good auditory retention, it also true that the acoustic homogeneity that confers these perceptual advantages is not very typical of vowels in natural speech. Thus, in addition to favoring an auditory mode of processing, isolated vowels, by their very unnaturalness, may discourage phonetic processing and, in extreme cases, lose their speechlike quality altogether.

It remains for us to mention some categorical perception studies that varied properties of vowels other than their phonetic quality. One such property is duration, which carries some distinctive phonetic information in English, but much more in certain other languages, such as Thai. Bastian and

Abramson (1964) created a continuum from /baat/ to /bat/ (meaningful words in Thai) by removing pitch pulses from the center of a natural token of /baat/. Oddity discrimination scores were quite continuous for both Thai and American listeners, showing no evidence of a phoneme boundary effect. These results were further confirmed in a vocal imitation task where the duration of the responses was found to be a nearly linear function of the durations of the stimuli. (Thai subjects did show a slight effect of categorization here, but since Bastian and Abramson did not dwell on it, it was probably nonsignificant.) We have already mentioned (Section 5.2.1) the study by Raphael (1972), who showed that variations in vowel duration are not categorically perceived even when they cue a consonantal distinction (final consonant voicing).

Another property of vowels that carries phonemic significance in many languages, but not in English, is their pitch contour. Thai, for example, has five distinctive tones. Abramson (1961) generated a synthetic continuum between two of these on the fixed carrier /naa/. ABX discrimination results provided some evidence for a phoneme boundary effect in Thai listeners, but the results rested on a comparison of Thai and American listeners, since stimulus problems prevented a direct interpretation of discrimination functions. A subsequent study by Chan, Chuang, and Wang (see Wang, 1976) found evidence of a category boundary effect for Chinese subjects listening to a continuum of Mandarin tones. The effect disappeared, however, after practice in ABX discrimination. Abramson (1979) re-investigated the issue using a new continuum of Thai tones that consisted simply of flat frequency contours varying in level. 4IAX discrimination of these stimuli by Thai listeners was entirely continuous. Taken together, these three studies suggest that moving pitch contours may elicit a tendency toward categorical perception while static frequency levels do not.

5.2.6. Summary

A brief summary is in order after reviewing so many different studies. It is evident that the large majority of experiments obtained results consistent with categorical perception. Thus, categorical perception is not only characteristic of stop consonants, but also of nasals and, to some lesser degree, of liquids, semivowels, and fricatives. The perception of liquids, semivowels, and fricatives is clearly less categorical than that of stops, and that of fricatives, at least, may become entirely continuous under certain conditions. Vowels, too, show a phoneme boundary effect in most conditions, and may even be perceived fairly categorically when embedded in context. Indeed, there are few experiments in the literature that present conclusive evidence for perfectly continuous discrimination of a speech continuum.

5.3. Perception of Nonspeech Stimuli

From the very beginnings of categorical perception research, the comparison of speech and nonspeech stimuli has been of central interest. Initially, the purpose of these comparisons was to determine whether categorical perception was due to "acquired similarity" of different sounds from the same category (in which case nonspeech discrimination should be easier than within-category speech discrimination), "acquired distinctiveness" of sounds from different categories (in which case between-category speech contrasts should be easier to discriminate than nonspeech), or both (e.g., Liberman, Harris, Eimas, Lisker, & Bastian, 1961). As interest in this issue faded (Mattingly et al., 1971), it was replaced by a search for possible psychoacoustic bases of linguistic category boundaries and discrimination peaks. This required nonspeech stimuli as similar as possible to the speech stimuli they were to be compared with, but sufficiently dissimilar so as not to elicit speech-like percepts. Finding the right balance between these two requirements has been a major (and, perhaps, insurmountable) methodological obstacle.

5.3.1. Perception of Continua Unrelated to Speech

In the early stages of categorical perception research, it was important to make sure that perception of simple nonspeech continua was really continuous in the standard categorical perception paradigm. It seemed possible, after all, that categorical perception was an artifact of the procedures used, which differed in certain respects from those of psychophysical research.

An appropriate comparison was undertaken by Eimas (1963). He included, along with vowel and stop consonant continua, a continuum of noise bursts varying in duration and a visual continuum of different levels of reflectance (Munsell grey scale). Both nonspeech continua were presented in labeling and ABX tests. The labels were "long" or "short" for the noises, and "light," "medium," or "dark" for the visual stimuli. While both nonspeech continua were consistently labeled by the subjects, discrimination was far better than predicted and quite continuous. Thus, discrimination of the nonspeech stimuli was clearly not limited by categorization but, since discrimination scores were at or near the ceiling, Eimas did not provide a strong test of whether labels can have any influence on nonspeech discrimination.

Indeed, Cross et al. (1965), employing a visual continuum of sectored circles, found results not unlike categorical perception. Their subjects were first trained to give verbal labels to the stimuli. A subsequent ABX discrimination test revealed a clear peak at the category boundary. However, discrimination of within-category contrasts was considerably better than predicted on the basis of labeling performance, so that the data showed only "a degree of categorical perception typical of vowels" (Studdert-Kennedy et al., 1970, p. 242), not of stop consonants. Unfortunately, two independent replications of the Cross et al. study failed to find similar effects. Liberman, Studdert-Kennedy, Harris, and Cooper (1965), in a detailed critique of Cross et al., reported they could not find any discrimination peaks, before or after categorization training. It may be countered that they provided less formal training and that discrimination performance was too high to reveal any peaks. However, a second, almost exact replication of Cross et al. by Parks et al. (1969) revealed no consistent category boundary effects and no influence of categorization training.

More recently, Pastore (1976) also reported a failure to obtain a discrimination peak at the "alternation" vs. "movement" boundary for the visual Phi phenomenon (two lights alternating at varying rates). However, Kopp and Udin (1969) and Kopp and Livermore (1973) found a clear discrimination peak (in ABX and same-different tasks, respectively) on a continuum of pure tones varying in frequency, following classification training. (See Vinegrad, 1972, for corresponding results in a magnitude scaling study.) Kopp and Livermore performed a signal detection analysis of their data and found that the discrimination peak was entirely due to response bias, so that an unbiased measure of sensitivity was constant across the whole continuum. This finding contrasts with Wood's (1976a, 1976b) similar analyses of stop consonant discrimination, which showed both bias and sensitivity changes contribute to the phoneme boundary effect (cf. also Elman, 1979; Popper, 1972).

Healy and Repp (1982) recently constructed a nonspeech continuum consisting of brief, steady-state, single-formant resonances varying in frequency (timbre). The stimuli were presented in same-different and labeling tasks whose order was counter-balanced. Prior labeling experience did not seem to have any effect on discrimination performance, which exhibited a peak at the category boundary.

The data just reviewed suggest that category labels may influence nonspeech discrimination under certain circumstances. We might expect these circumstances to be those that make it difficult to rely on auditory memory--that is, when the differences to be detected are small to begin with. A role for some form of categorical encoding in discrimination is also predicted by the psychophysical dual-coding theory of Durlach and Braida (1969). In all nonspeech studies mentioned, however, within-category discrimination was substantially better than predicted by the Haskins model; perception was never truly categorical.

The studies discussed so far looked for category boundary effects on obviously continuous physical dimensions; therefore, if such effects were found, they must have been due either to response bias introduced by the subjects' category labels or to procedural artifacts. On the other hand, some recent studies have demonstrated category boundary effects on continua that straddle a psychophysical threshold. These findings are hardly surprising; the point of these studies was, however, to lend plausibility to the hypothesis that boundary effects on speech continua might likewise be caused by psychophysical discontinuities, not by categorization per se.

Some pertinent data were reported by Pastore et al. (1977). In one experiment, they flashed a light at various rates centered around the flicker fusion threshold. The subjects were able to label the stimuli consistently as "flicker" or "fusion," and ABX discrimination results showed a peak at the boundary and poor discriminability within categories. In a second experiment intended to have some relevance to speech perception, Pastore et al. varied the intensity of a pure tone that alternated with a constant reference tone of the same frequency. ABX discrimination scores showed a peak at the boundary between the two (arbitrary) categories used by subjects in the labeling task. In a control condition, the reference tone was omitted, and the discrimination peak disappeared. Pastore et al. mention, however, that they failed to replicate these results using noise stimuli, and their data for tones seem fairly variable. For these reasons, the claim of Pastore et al. that a fixed reference stimulus generates a sharp boundary and a corresponding discrimination peak must be accepted with caution. It is also clear from their discussion that good within-category discrimination would have been possible at larger step sizes, so that perception was not truly categorical.

In all the cases discussed in this subsection, the categories were not particularly familiar, sometimes even arbitrary. This is also true for the majority of the various nonspeech analogs of speech, to be discussed next. However, there are also nonspeech domains associated with highly overlearned categories; two of them (color and music) will be considered in the final subsection (5.3.5).

5.3.2. Nonspeech Analogs of Voice Onset Time

The primary cue for the voicing distinction in initial stop consonants is temporal--the delay of the onset of voicing relative to the stop release. On the positive (voicing lag) side, this temporal delay results in correlated spectral changes: The interval prior to voicing onset is filled with aperiodic noise (except in the earliest studies where only "F1 cutback" was manipulated), there is no energy in the region of the first formant before the onset of voicing, and at voicing onset the formants (F1 in particular) start at frequencies close to those of the following vocalic portion. These spectral correlates of voice onset time (VOT) all are relevant to the perception of the voicing distinction, but most studies have focused on the temporal aspect of VOT only.

The first attempt to devise nonspeech analogs of VOT was undertaken by Liberman, Harris, Kinney, and Lane (1961). They synthesized a /do/-/to/ continuum by delaying the onset of F1 in varying amounts. A matched nonspeech continuum was obtained by playing the stimuli with the frequency scale inverted, so that F1 was in the region previously occupied by F3, and vice versa. (This was literally possible on the Haskins Laboratories Pattern Playback.) In addition, the initial transition of the new F1 (previouly F3) was modified, to assure that the stimuli would not sound speechlike. While ABX discrimination of the speech stimuli was highly categorical, that of the nonspeech stimuli was extremely poor and barely exceeded chance even at the at the largest step size used. In other words, speech discrimination was vastly superior to nonspeech discrimination. Liberman et al. interpreted this finding as evidence for the acquired distinctiveness (rather than acquired similarity) of speech sounds. They did acknowledge, however, that there were a number of differences between speech and nonspeech stimuli, which may have been responsible for the poor performance with the latter.

Liberman et al. did not ask their subjects to label the nonspeech stimuli. Lane and Schneider (1963; cited in Lane, 1965) found that some subjects could be trained to label them as accurately as the speech stimuli. In a subsequent ABX test, these subjects produced above-chance discrimination scores with a peak at the boundary. This report was questioned, however, by Studdert-Kennedy et al. (1970), whose detailed examination of the Lane and Schneider data revealed that they were extremely variable and hardly conclusive. Studdert-Kennedy et al. also reported a failure to replicate the results with five subjects, none of whom could be trained to label the nonspeech stimuli in a consistent way.

The /do/-/to/ control stimuli may have been too complex for listeners to detect the relevant differences without extensive training. Later studies used stimuli of a simpler acoustic structure. Hirsh's (1959) finding of a threshold in the vicinity of 20 msec for determining the temporal order of two auditory events stimulated the thought (Liberman, Harris, Kinney, & Lane, 1961) that this threshold might be related to the category boundary on a VOT continuum. This suggestion makes good sense when applied to speech stimuli generated by the method of F1 cutback, where the onset of low-frequency energy may indeed either precede or follow the onset of high-frequency energy. However, it loses some of its appeal when aspiration enters the scene (as it does in more sophisticated--and more appropriate--VOT synthesis), for aspiration always precedes the onset of voicing and provides a powerful cue to the voicing distinction. It has also been long known that VOT boundaries tend to be at rather longer onset asynchronies (especially for alveolar and velar stops) than the temporal-order threshold (Lisker & Abramson, 1970). Nonetheless, a good deal of research has been generated by this presumed analogy.

Stevens and Klatt (1974) synthesized stimuli consisting of a 5-msec broadband noise burst followed by a variable silent interval and steady-state formants roughly appropriate for the vowel /ɛ/. According to these authors, "none of the stimuli could be readily interpreted as speech events" (p. 654). Listeners were asked to label the stimuli according to whether or not they heard a silent interval between the noise and the vowel. The category boundary fell at about 20 msec of "voice onset time" (measured from the onset of the burst), which matched the time obtained by Hirsh (1959) with tones. However, no discrimination data were obtained for these stimuli, and their analogy to VOT in speech may be questioned because of the absence of aspiration noise. Their relation to Hirsh's findings is equally doubtful, for the task did not require temporal order judgments but detection of a gap.

These objections do not apply equally to a subsequent study by Miller et al. (1976), who presented white noise and a square-wave buzz at varying noise-buzz lead times in labeling ("no-noise" vs. "noise") and oddity discrimination tasks. The listeners were experienced in psychoacoustic experiments. Their category boundaries varied widely (from 4 to 31 msec of noise lead time), but they showed clear discrimination peaks, which in all cases but one coincided with the boundary. Control results obtained with isolated noises did not reveal any discrimination peaks. Miller et al. compared their results with those of Abramson and Lisker (1970) for VOT and found a striking similarity of the average discrimination functions. However, they neglected to point out that at least three of their eight listeners had category boundaries at substantially shorter values of noise lead time (4-8 msec) than are ever obtained with speech stimuli varying in VOT. Such a wide range of individual differences in boundary locations is quite atypical of speech and presumably reflects variations in auditory acuity or response criteria, since all listeners were quite experienced. Therefore, while Miller et al. have shown (as have Pastore et al., 1977) that results resembling categorical perception can be obtained with nonspeech stimuli straddling a psychophysical threshold, they have not presented a convincing case for any direct correspondence of the category boundaries in speech and nonspeech.

Of course, it could always be argued that the supposed nonspeech analogs of VOT simply fell short of the mark. As we pointed out above, if the analogs are made too speechlike, there is the danger that they are perceived as speech. Wood (1976a) accepted this risk when he decided simply to excise most of the steady-state vowels of stimuli from a /ba/-/pa/ continuum (ranging from -50 to +70 msec of VOT) and to use the initial 120 msec as "nonspeech analogs." According to Wood, who interviewed his subjects carefully, these truncated stimuli were not spontaneously categorized as (or even recognized as being related to) speech. (They were not presented for identification at all.) Same-different discrimination results for full and truncated syllables were similar at short VOTs, but at long VOTs the scores for the truncated stimuli were rather high, which obscured the discrimination peak that may otherwise have been obtained. Most likely, the reduction in the duration of the periodic portion with increasing VOT became detectable at long VOTs in the truncated stimuli. Wood also mentions that identical results were obtained in a subsequent unpublished experiment, where subjects were instructed to hear the short syllables either as speech or as nonspeech. He concluded that "the phoneme boundary effect for VOT does not depend exclusively upon phonetic categorization but may reflect acoustic and auditory properties which are independent of phonetic processing" (p. 1388). Unfortunately, Wood's results cannot be considered conclusive because of the confounding of VOT with "vowel duration" in the truncated stimuli.

Following a previous unpublished study by Ades (1973), Pisoni (1977) employed a temporal order judgment task to examine how much it might have in common with VOT perception (cf. also Pastore, Harris, & Kaplan, 1982). He varied the relative onset times of two pure tones similar in frequency to F1 and F2 of a neutral vowel, and trained subjects to classify these stimuli into two categories exemplified by the extreme (50 msec) low-tone lead and lag stimuli. As it happened, the category boundary of most subjects fell not at the point of simultaneous onset but at short low-tone lags (where, accepting the analogy with F1 cutback, the VOT boundary is located). Discrimination peaks at the subjects' boundaries were obtained in a subsequent ABX task with feedback. In a second experiment, the ABX test was presented without prior training in labeling. Some subjects showed results similar to the first experiment, while others showed two discrimination peaks, at approximately 20-msec lead and lag times of the lower tone. The double peaks suggested that there were two "natural boundaries" on the continuum, one corresponding to the detection threshold for low-tone leads and the other, to that for low-tone lags. This hypothesis was strengthened by a further experiment in which subjects were successfully taught to classify the stimuli into three categories.

Pisoni concluded on the basis of these data that a "basic limitation on the ability to process temporal-order information" (p. 1360) underlies the perception of VOT, acknowledging at the same time that the location of the voicing boundary is influenced by a variety of other factors, ranging from spectral signal properties to the subjects' linguistic background (cf. Section 6.2). However, Pisoni's conclusion provides, at best, an incomplete account of VOT perception, for the voiced/voiceless distinction for syllable-initial stops in English rests as much on the perceived presence of aspiration or of a high F1 onset as on the temporal cue of delay of voicing onset. Also, it is not clear how factors such as linguistic experience might modify the location of a strictly psychoacoustic boundary. It seems more likely that psychoacoustic and linguistic boundaries coexist.

That the tone-onset-time (TOT) continuum used by Pisoni is not a very close analog of VOT is suggested by several recent findings. Pisoni (1980a) himself failed to find a selective adaptation effect of TOT stimuli on syllables from a VOT continuum or vice versa, which suggests that the two types of stimuli do not engage the same auditory mechanisms. Rather convincing evidence for a fundamental difference between VOT and TOT was obtained by Summerfield (in press), who used, in addition, noise-buzz stimuli similar to those of Miller et al. (1976). All three sets of stimuli were constituted of two steady-state components analogous to F1 and F2 and closely matched in frequency and amplitude across the three sets. Summerfield investigated the influence of the frequency of the lower-frequency component (F1 or its analog) on the location of the boundary. On the VOT continuum (labeled "g" or "k"), he found, in accordance with previous results (Summerfield & Haggard, 1977), a shift of the boundary toward longer values as F1 frequency was raised. However, there were no comparable effects on the two nonspeech continua (labeled "simultaneous onset" or "successive onset"). Even granting that the use of phonetic labels for the speech stimuli only may have contributed to the difference, these results seriously weaken the proposal that the VOT boundary is merely a temporal-order threshold (or even, for that matter, a noise-detection threshold).

It appears, however, that the last word on this issue has not yet been spoken. Hillenbrand (1982) recently reported an effect of the duration of a simulated F1 transition on the TOT boundary. Although the details of this study are not available at this time, it seems possible that Hillenbrand's stimuli, which contained frequency transitions in both tones, were sufficiently speechlike to elicit a phonetic mode of processing (cf. Grunke & Pisoni, 1979; Schwab, 1981). We might also take note of Molfese's (1978, 1980) analysis of evoked potentials to VOT and TOT stimuli. For both kinds of stimuli, a right-hemisphere component was found that distinguished between short-lag and long-lag stimuli, and also between different extents of long lags but not of short lags. This component seems consistent with a temporal-order threshold. It is evident that the question about the psychoacoustic bases of VOT perception is far from resolved.

5.3.3. Nonspeech Analogs of Formant Transition Cues

The critical cues for distinguishing different places of articulation in synthetic stop consonant continua are the transitions of F2 and F3. In the earliest continua, only two formants (F1 and F2) were used. This suggested an obvious nonspeech control: to omit the constant signal portions (F1, and perhaps also the steady state of F2) and to present F2 (or only the F2 transition) by itself. Several studies have investigated the perception of these isolated transitions ("chirps") or transitions plus steady state ("bleats"). It should be noted that while chirps sound rather nonspeechlike, they may be associated with speech sounds when subjects are provided with appropriate labels (Nusbaum, Schwab, & Sawusch, 1981). Bleats have some resemblance to strongly nasalized stop-vowel syllables and therefore are problematic as a nonspeech control. Studies employing these stimuli, however, invariably report that naive listeners do not perceive them as speech.

Kirstein (1966) was the first to present bleats in an ABX discrimination task. These isolated second formants were derived from the two-formant /be/-/de/-/ge/ continuum of Liberman et al. (1957) by omitting the constant F1. While the speech stimuli had been discriminated fairly well (at the level predicted by the Haskins model or better), discrimination of the bleats was at chance at all step sizes used. However, when the bleats were played backwards, so that the transition was at the end, discrimination was better than chance and improved as step size increased.

A more comprehensive study along the same lines was conducted by Mattingly et al. (1971). They used both bleats and chirps, derived from continua of initial and final stops. Oddity discrimination scores for chirps and bleats were rather similar and noncategorical, and discrimination was easier when the transitions were at the end (more precisely, when offset frequencies varied, rather than onset frequencies), which confirmed Kirstein's results and was in agreement with existing psychophysical data (Brady et al., 1961). Due to peaks in the boundary regions, discrimination of syllable-initial stops was superior to discrimination of the corresponding nonspeech stimuli. The relationship was reversed for syllable-final stops whose discrimination function was also more similar to those for the corresponding nonspeech stimuli. However, Popper (1972) employed F2 bleats with final transitions and three-formant vowel-consonant syllables and found that, while the overall discriminability of speech and nonspeech was similar, the speech discrimination function showed a broad peak at the boundary while the nonspeech function did not.

In another related study, Syrdal-Lasky (1978) presented F2 chirps in an oddity discrimination task at three different intensities. While, at the two higher intensities, the discrimination functions were nearly flat, at the lowest intensity there were two discrimination peaks. The peaks resembled those obtained with a simple /pæ/-/tæ/-/kæ/ continuum consisting of the chirps followed by a steady-state F1-F2 pattern. These data deserve to be replicated, for they are the only instance so far of boundary effects on a chirp continuum.

Pisoni (1971: Exp. II) used bleats with initial transitions as stimuli in a training experiment, intended to test Lane's (1965) proposition that categorical perception of nonspeech stimuli could be acquired in the laboratory. The stimuli were derived from a /bæ/-/dæ/ continuum, and listeners were given these labels to use. Although training did improve both labeling consistency and discrimination accuracy, there was no evidence that it introduced any consistent phoneme boundary effects. Moreover, discrimination following training was generally much better than predicted by the Haskins model, suggesting noncategorical perception. In a later replication, however, Pisoni (1976b) obtained not only very steep labeling functions but also discrimination peaks at the category boundary for most listeners. It is not clear what caused this difference in results. Pisoni (1976b) states only that his earlier study was "not entirely satisfactory for a number of reasons" (p. 125), and he does not discuss the possibility that the bleats were heard as speech (/mæ/-/næ/) by the subjects. However, that possibility seems very real, and one is led to wonder whether the same results would have been obtained, had arbitrary labels been used, or the same labels in reverse assignment.

Isolated F3 resonances were presented in two studies of the /r-l/ contrast (McGovern & Strange, 1977; Miyawaki et al., 1975). Although located at higher frequencies than F2 bleats derived from stop consonant continua, they are easier to discriminate because they have a distinctive steady state and slower transitions. As with bleats, however, discrimination is easier when the distinctive information is located at the end (as it is in vowel-liquid stimuli) than when it occurs at the beginning (McGovern & Strange, 1977). In both studies cited, F3 discrimination results showed no resemblance to /r/-/l/ discrimination.

So far, there is no convincing evidence that chirps or bleats yield a "boundary effect" when they are perceived as nonspeech. To avoid the objection that chirps and bleats are poor analogs of speech because so much of the original acoustic context (F1, F3) has been removed, Bailey, Summerfield, and Dorman (1977) constructed "sine-wave analogs" of speech stimuli: The first three formants of /bo/-/do/ and /be/-/de/ continua were mimicked by three pure tones (cf. Cutting, 1974). The interesting fact about sine-wave analogs is that they may be heard as speech with experience or appropriate instructions, but sound like nonspeech whistles to naive subjects. (While this is also true, to some extent, for chirps and bleats, the phonetic and nonphonetic interpretations of sine-wave analogs appear to be more disparate in the listener's experience, which makes introspections a reliable source of information about perceptual modes.) Bailey et al. presented their speech and nonspeech stimuli in AXB identification (i.e., classification without labels) and discrimination tasks. The sine-wave stimuli were presented twice, first without and then with instructions to hear them as speech. The speech continua had been chosen to yield boundaries in different locations, one to the left and one to the right of the center of the stimulus range. Although classification accuracy was not very high, the expected difference in boundaries was obtained for the speech stimuli as well as for the sine-wave stimuli under speech instructions. However, under nonspeech instructions the boundaries on the two continua coincided in the center of the stimulus range. The discrimination functions for the two sine-wave continua showed corresponding differences in the speech condition, but no difference in the nonspeech condition. Unfortunately, the discrimination scores were rather low and did not show pronounced peaks, probably due to the poor labeling performance. In a second experiment, Bailey et al. used a /ba/-/da/ continuum and its sine-wave analog and divided subjects into speech and nonspeech groups on the basis of post-experimental interviews. Again, the category boundary on the sine-wave continuum resembled that on the speech continuum when the sine-wave stimuli were heard as speech, but not when they were heard as nonspeech.

The significant work of Bailey et al. has remained unpublished and still awaits replication, particularly as far as the discrimination results are concerned. Together with the earlier chirp and bleat data, however, it strongly suggests that the location of the category boundary as well as the shape of the discrimination function are not determined by acoustic stimulus properties alone. The contribution of Bailey et al. lies, in part, in their attention to listeners' introspections as an indicator of perceptual modes. Pisoni (1976a), in an interesting pilot study, may have failed to take this aspect into consideration. He synthesized sine-wave analogs of a /ba/-/da/-/ga/ continuum, omitting the steady-state portion, so that only the initial 50-msec transitions remained. Three experienced listeners generated ABX discrimination functions that exhibited two peaks, approximately where the phoneme boundaries would lie on the corresponding speech continuum. Pisoni took this as support for the hypothesis that psychoacoustic discontinuities related to phonetic boundaries existed on the sine-wave transition continuum. However, in view of recent demonstrations that initial formant transitions without a following steady-state vowel can be quite accurately labeled as stop consonants (Blumstein & Stevens, 1980; Jusczyk, Smith, & Murphy, 1981; Tartter, 1981), it seems not impossible that Pisoni's experienced listeners were able to achieve this also with the sine-wave analogs.

However, Pisoni's (1976a) results receive support from another unpublished study (Wood, 1976b). Wood presented the initial 40 msec of synthetic stimuli from a /bæ/-/dæ/-/gæ/ continuum in a same-different task and obtained clear indications of increased perceptual sensitivity (in terms of a bias-free measure) at the points where the category boundaries for the full syllables were located. Significantly, Wood interviewed his subjects very carefully and determined that they did not relate the truncated stimuli in any way to the full syllables. The plausibility of this finding is increased by a comparison of Wood's results with Tartter's (1981): Using similar stimuli under speech instructions, Tartter obtained better discrimination performance for truncated than for full syllables, while Wood obtained the opposite, suggesting that Wood's subjects indeed did not hear the stimuli as speech. (However, Wood goes on to mention that, in a subsequent study, he did not find any effect of instructions, which is puzzling.)

Given the excellent reputation of both Pisoni and Wood as careful researchers, their findings may be taken as highly suggestive of psychoacoustic boundaries on a place-of-articulation continuum. However, it is difficult to reach a firm conclusion on the basis of unpublished and partially conflicting (Bailey et al., 1977) evidence.

5.3.4. Nonspeech Analogs of Closure Cues

Nonspeech analogs of the closure duration cue for intervocalic stop voicing were constructed by Liberman, Harris, Eimas, Lisker, and Bastian, (1961). The stimuli consisted of two noise bursts whose durations (about 200 and 80 msec) and amplitude envelopes matched those of the pre- and postclosure portions of speech stimuli (/ræbId/-/ræpId/), and which were separated by varying intervals of silence (30-120 msec). ABX discrimination of silence in this nonspeech context was consistently inferior to its discrimination in speech context, and there were no pronounced peaks in performance. At the time, these results were welcomed as support for the "acquired distinctiveness" hypothesis. Further support came from a study by Baumrin (1974), who found, in an information-theoretic analysis, that less information was transmitted on a nonspeech continuum of silence durations than on a corresponding speech continuum.

Perey and Pisoni (1980) recently examined the discrimination of silence embedded between two 250-msec three-tone complexes (imitating the first three formants of /ə/-like vowels) with or without simulated formant transitions into and out of the closure. Even though the subjects were first taught to classify the stimuli into two categories, subsequent ABX discrimination was extremely poor and entirely continuous. Although both this study and that of Liberman et al. (1961) suffered from a (somewhat unnecessary) floor effect, they certainly demonstrated striking differences in listeners' sensitivity to silence duration in and out of speech context.

Silence is also an important cue for stop manner. A second cue in prevocalic position is a rapidly rising F1 transition. These two cues can be traded off against each other, within limits: For example, less silence is needed to hear "stay" rather than "say" when the onset of F1 in the vocalic portion is low than when it is high. Best et al. (1981) examined whether this trading relation is found in sine-wave analogs of "say"-"stay" stimuli, consisting of an initial noise burst followed by a variable silent interval and a three-tone complex with variable onset frequency of the lowest (F1-analog) tone. The results of labeling and oddity discrimination tasks provided a positive answer, but only for those subjects who reported that they perceived the sine-wave stimuli as speech. The remaining subjects, who reported various nonspeech impressions, fell into two groups--those that appeared to pay attention to the temporal cue (gap duration) and those that paid attention to the spectral cue (onset quality of the simulated vocalic portion). The discrimination results for these two groups differed radically: The scores of the temporal listeners were somewhat lower than those of the speech listeners and exhibited two unpredicted peaks (at about 20 and 65 msec of silence, respectively) that warrant further investigation. The scores of the spectral listeners, on the other hand, were extremely high and much superior to those of the speech listeners. Those listeners who interpreted the stimuli as speech adopted neither of these selective-attention strategies but instead seemed to integrate the two cues into a single (phonetic) percept that, as the comparison with the nonspeech listeners shows, at the same time aided and hindered discrimination. These findings of Best et al. provide some of the most convincing evidence for the existence of separate modes of perception for speech and nonspeech.

To provide a potential nonspeech analog for the fricative-affricate contrast, one important cue for which is amplitude rise-time, Cutting and Rosner (1974, 1976) varied the rise times of tonal stimuli (sawtooth or sine waves). These stimuli had the special distinction of conveying a manner contrast important in music, "pluck" vs. "bow." Thus, unlike any of the other nonspeech controls discussed so far, these stimuli spanned two natural musical categories. Comparing affricate-fricative (/tʃa/-/ʃa/, /tʃæ/-/ʃæ/) and pluck-bow continua in standard identification and discrimination tasks, Cutting and Rosner found categorical perception for both. This result suggested, more than any other, that a speech contrast had been built on a pre-existing auditory threshold, and it became one of the most widely cited and replicated findings of recent years (e.g., Cutting, 1978; Cutting et al., 1976; Jusczyk, Rosner, Cutting, Foard, & Smith, 1977; Remez, Cutting, & Studdert-Kennedy, 1980). All replications, however, used the original pluck-bow stimuli provided by Cutting and Rosner. It was embarrassing, therefore, when Rosen and Howell (1981) analyzed these stimuli and found them to be not equally spaced along the rise-time continuum. They conducted a series of very careful experiments and failed to find categorical perception with equally-spaced stimuli; on the whole, rise-time discrimination followed Weber's law, and there was no effect of prior labeling experience. These results were replicated by Kewley-Port and Pisoni (1982). It thus appears that the findings of Cutting and his colleagues must be dismissed as artifactual.

In summary, despite a few suggestive results, there is no conclusive evidence so far for any significant parallelism in the perception of speech and nonspeech. What seems to matter is not whether the stimuli *are* speech or nonspeech but how listeners interpret ("hear") them (see also Section 6.1). Categorical perception appears to be a function not so much of the physical properties of the stimuli as of the frame of reference adopted by a listener.

5.3.5. Categorical Perception of Color and Music

A brief excursion is in order into domains that, like speech, employ highly overlearned categories. Here the question arises, as it does for speech, whether the category distinctions have a psychophysical basis or whether they are essentially arbitrary and determined by cultural convention. While the role of cultural factors and experience in speech perception will be discussed in Section 6.2, we will touch on these topics as we discuss briefly some relevant findings on color and music perception.

To determine whether color discrimination performance covaries with color categorization, Lane (1967) compared data from earlier color labeling and discrimination studies and discovered that discrimination performance indeed showed peaks at the boundaries between the major categories (violet, blue, green, yellow, red). This finding was replicated by Kopp and Lane (1968) with two American subjects and compared to data obtained from two speakers of a Mexican Indian language (Tzotzil) whose color categories divide the wavelength continuum in a different fashion. Kopp and Lane interpreted their data as showing an influence of linguistic habits on discrimination, but a look at their figures makes their conclusion seem unwarranted. To the extent that one can conclude anything from comparing groups of two subjects each, the discrimination functions of American and Tzotzil subjects seemed not fundamentally different. There appears to be little other evidence in favor of Kopp and Lane's thesis in the literature; on the contrary, there are studies showing that linguistic habits have no influence on the accuracy of color discrimination (Heider & Olivier, 1972). This suggests that the peaks in the color discrimination function have a psychophysical rather than a cultural basis.

Further support for this hypothesis comes from studies of color discrimination in infants. Using a habituation procedure, Bornstein, Kessen, and Weiskopf (1976) found that 4-month-old infants were more sensitive to hue differences across (adult) category boundaries than within categories. There is also anthropological evidence that the basic color categories are similar throughout the world, although some cultures use more different categories than others (Berlin & Kay, 1969). All this ties in with extensive physiological evidence for two opponent-process mechanisms in the neural coding of color, so that the peaks in color discrimination are likely to have a direct physiological explanation. Bornstein (1973) has even proposed that certain cross-cultural differences in color naming can be explained by known racial variations in visual anatomy. We should mention that color perception was never a serious candidate for true categorical perception, for although it shows discontinuities in discrimination, many different hues can be distinguished within color categories. Color perception exhibits a category boundary effect, but it is far from categorical.

Results closer to true categorical perception have been obtained with musical stimuli. Musicians encounter a variety of explicit or implicit categories relating to intervals, chords, scales, timbres, attacks, etc. The ill-fated research on the pluck-bow distinction (Cutting & Rosner, 1974) has been mentioned above; this contrast, at least, does not seem to be categorically perceived. Most other research has been concerned with musical intervals (i.e., successive tones) or chords (i.e., simultaneous tones). One interesting aspect of music perception research is that familiarity with the distinctions involved varies enormously in the general population. Unlike speech, musical stimuli do not "name themselves." Comparisons of practicing musicians with "nonmusicians" provides information similar to that gained from comparing speech with nonspeech controls. (This author knows of no experiments conducted outside the reaches of traditional Western music.)

Siegel and Siegel (1977a) showed that musicians can accurately label intervals drawn from a continuum ranging from unison to a major triad, while nonmusicians exhibit very inconsistent labeling performance. In a subsequent study, Siegel and Siegel (1977b) obtained musicians' magnitude estimates for intervals ranging from a fourth to a fifth. They obtained plateaus and reduced variability within the three interval categories (fourth, tritone, fifth), and rapid changes with high variability at the boundaries. This suggested categorical perception, although no standard discrimination test was administered.

The classical methods of assessing categorical perception were applied to musical intervals by Burns and Ward (1978). They presented intervals ranging from a major second to a tritone in labeling and two-interval forced-choice (2IFC) tasks. (The pitch of the first note of each interval varied randomly.) The discrimination functions were strongly categorical and closely matched the predictions generated by the Haskins model, although within-category discrimination was somewhat better than predicted. Varying the interstimulus interval between two successive intervals from 300 msec to 3 sec, they did not find any change in performance, which is reminiscent of the similar (near-)absence of an effect of temporal delay with stop consonants (Pisoni, 1973). Subsequently, Burns and Ward determined 2IFC difference limens, using a staircase method and testing their subjects until they reached asymptote. The results showed improved and more nearly continuous discrimination. The discrimination performance of a group of musically untrained subjects was much poorer but essentially continuous, which led Burns and Ward to conclude that musical intervals are learned, not natural, categories.

The categorical perception of simultaneous intervals or chords was first investigated by Locke and Kellar (1973). They presented chords consisting of three tones, with the frequency of the middle tone varying. The chords spanned the range from a minor triad to a major triad, but the subjects were not provided with these labels and instead classified the stimuli by matching them to a standard (one of the two endpoint stimuli). There was considerable individual variability, and non-musicians' performance was very poor. Musicians, on the other hand, showed a clear category boundary together with pronounced peaks in same-different discrimination scores; within-category discrimination, however, was much higher than predicted. A closer fit between predicted and obtained scores was obtained by Blechner (1977), who presented chords from a minor-major continuum in standard labeling and oddity discrimination tasks. Those subjects who were able to label the stimuli consistently as "minor" or "major" also showed fairly categorical discrimination, although scores were somewhat higher than predicted. A number of subjects were unable to label the chords consistently; their discrimination scores were low and showed no peak. Blechner also included a control consisting of only the middle tones of the chords. These stimuli were identified without difficulty as "low" or "high" by all subjects and discrimination performance was noncategorical, though higher for trained musicians. Zatorre and Halpern (1979) essentially replicated Blechner's results for chords, using two-tone simultaneous intervals (from minor third to major third).

Categorical perception of stimuli varying in rhythm was reported by Raz and Brandt (1977). The stimuli consisted of three consecutive tones, with the temporal position of the second tone varying. Since only an abstract of their study is available, it is not clear how categorical the results really were.

In summary, the musical results contrast with the color results--apart from the difference in modality--in that the former seem to reflect learned categories while the latter reflect natural, physiologically based categories. While category boundary effects are obtained in either case, perception is (interestingly) more nearly categorical in the case of the learned categories. Of course, their acquiredness does not necessarily mean that they do not have a physical basis: Musicians may learn to discover acoustic categories (e.g., simple frequency ratios) that simply are not registered by nonmusicians. Still, the fact that these categories must be established through experience, and that they have an effect in perception once they have been learned, is highly relevant to our understanding of speech perception. Specifically, it supports the hypothesis that categorical perception of speech is a product of categories acquired in the context of a particular language, and not of prewired psychoacoustic sensitivities (see Section 6.2).

6. SUBJECT FACTORS IN CATEGORICAL PERCEPTION

In this section we will consider the contribution that the listener makes to categorical perception. Here we will encounter evidence that is of vital importance to understanding the phenomenon. In Section 6.1, we will first review the effects of experience and extensive practice on speech discrimination, as well as the roles played by expectations and strategies. Section 6.2 discusses the important and rapidly expanding research comparing listeners of different language backgrounds or attempting to teach unfamiliar phonetic distinctions to subjects. Section 6.3 briefly comments on infant speech perception. While this research is of prime importance, a detailed review will not be provided here, as several excellent and comprehensive discussions have recently appeared in the literature. In the final subsection, 6.4, the topic will be the small and somewhat controversial literature on categorical perception in nonhuman animals.

6.1. Practice and Strategies

6.1.2. Effects of Discrimination Training

In Sections 4.2.1 and 5.1.1, we have reviewed several studies showing that within-category discrimination on a stop consonant continuum can be improved somewhat by using more sensitive discrimination paradigms, such as 4IAX (e.g., Pisoni & Lazarus, 1974). One of the largest increases in discrimination performance was obtained by Hanson (1977), who provided feedback throughout a same-different reaction-time task, together with careful instructions to detect physical differences between stimuli (which contrasted with phonetic matching instructions in a second condition). The effectiveness of feedback is illustrated by a comparison of Hanson's results with those of Repp (1975), who used essentially the same task and instructions but did not provide any feedback: His subjects failed to show any improvement.

The exact role of instructions on the degree of categorical perception is not quite clear. It is possible that inexperienced subjects do not always understand the meaning of "physical differences" among speech sounds, and some excessively categorical results in the literature may reflect that fact. What is more likely is that naive subjects do not know what sort of physical difference to listen for (see Pastore, 1981; Pisoni, 1980b). Some training with feedback may be necessary to direct their attention to the relevant auditory qualities, which are often difficult to convey by instructions alone.

Another procedural change that seems to improve performance is to restrict the discrimination task (or part of it) to within-category comparisons only. The mixing of between- and within-category contrasts in the same block of trials, which has been the standard procedure in nearly all the studies reviewed so far, may place an attentional burden on the subjects that prevents them from focusing effectively on nonphonetic stimulus attributes. In addition to biasing subjects toward using a phonetic criterion, this mixing of different stimulus comparisons increases "subject uncertainty," which is known to increase psychophysical discrimination thresholds (Pastore, 1981).

A first attempt to improve VOT discrimination through extensive training was undertaken by Strange (1972). However, although she provided feedback, she used the standard oddity paradigm and a wide range of stimuli, which may have hindered her purpose. After a number of training sessions, discrimination performance had improved only slightly, primarily in the region of short voicing lags. A shift of labeling boundaries to shorter VOTs was also noted, which may account for the changes in discrimination performance. Although this shift may itself be taken to indicate an increased sensitivity to voicing lags, Strange's training study was considered unsuccessful both by herself and by later authors (Pisoni, Aslin, Perey, & Hennessy, 1982). It seems likely that the high-uncertainty discrimination paradigm prevented the accurate detection of acoustic differences (see also Section 6.2.2).

A fixed-standard AX task without feedback or extensive training was recently used by Repp (1981b) to assess the discriminability of within-category differences on several different speech continua. He found rather good performance on continua that varied silence duration ("say"-"stay," "say shop"-"say chop") but poor discrimination of VOT within the voiceless stop category. Repp (1981c), using the same paradigm, also found poor and seemingly categorical discrimination of fricative-vowel syllables by naive subjects. Thus, without training and/or feedback, low-uncertainty tasks do not lead to a dramatic improvement in discrimination performance. The secret lies in combining these procedures.

A fixed-standard AX task with feedback, using only two different stimuli in a whole block of trials, was employed first by Sachs and Grant (1976), who determined difference limens (d' = 1) on a /ga/-/ka/ VOT continuum. They reported threshold values of less than 2 msec with a 10-msec-VOT standard, and of 10 msec with a 60-msec standard, which clearly is far superior to any within-category performance obtained in previous studies. Also, the magnitude of the threshold increased monotonically with the VOT of the standard; that is, there was no phoneme boundary effect--a somewhat atypical result that was perhaps due to the use of subjects that were highly experienced in psychoacoustic tasks.

Ganong (1977) used a similar procedure to determine the discriminability of 15-msec VOT differences within the /pa/ category of a /ba/-/pa/ continuum. He found d' scores close to 1.0, which is obviously better than chance, although not quite as good as the Sachs and Grant difference limens for experienced subjects. Interestingly, Ganong's subjects were equally accurate (following AX discrimination training) in an absolute identification task in which the standard and comparison stimuli were presented singly and randomly, separated by several seconds. Thus, it appears that the subjects eventually achieved discrimination not by physically comparing the stimuli but by referring to some long-term internal representations.

A third study using the fixed-standard AX procedure (and the first to be published) was conducted by Carney et al. (1977). These authors paired all stimuli from a /ba/-/pa/ continuum (including negative as well as positive VOTs) with selected standards and obtained discrimination functions before and after extensive training with feedback. A conventional oddity discrimination task was also administered. In both discrimination tasks, performance was fairly categorical before training but vastly improved after training. Discrimination was still best in the category boundary region, but secondary peaks emerged within categories, particularly around 20 msec of prevoicing--a little-noted finding that is in accord with Pisoni's (1977) results for tone onset times. Phonetic labeling remained unaffected by training, and discrimination accuracy was equally high when subjects were required to provide labels following each "same-different" response. Finally, the trained subjects were even able to establish a new, arbitrary category boundary (at -50 msec of VOT) through identification training with feedback.

In a continuation of the research of Carney et al., Edman, Soli, and Widin (1978) observed that subjects trained on a labial VOT continuum could transfer their discrimination skills without any loss to a velar VOT continuum, and vice versa (see also Edman, 1979). However, discrimination remained most accurate in the boundary regions of both continua. In an application of the same techniques to place-of-articulation continua, Edman (1979) trained subjects on either a /bæ/-/dæ/-/gæ/ or a /pæ/-/tæ/-/kæ/ continuum and obtained excellent within-category discrimination and almost complete transfer to the other stimulus series.

Samuel (1977) demonstrated that a substantial improvement in within-category discrimination on a VOT continuum (/da/-/ta/, positive VOTs only) may also be obtained by training subjects in the ABX format, given that a fixed standard and feedback are provided. The performance increase occurred primarily in the /da/ category, suggesting that discrimination of very short voicing lags was not limited by a simultaneity/successiveness threshold. A discrimination peak at the category boundary remained, which Samuel ascribed to phonetic categorization. By espousing a two-factor model, Samuel contrasts with Carney et al., who favor a single-factor view, ascribing the boundary effect to psychoacoustic factors.

Several other training studies will be discussed in Section 6.2, since they were concerned more with establishing a new phonetic contrast than with improving within-category discrimination. We have also omitted from discussion several studies that tested adults in low-uncertainty paradigms to provide comparison data for infants or animals run under the same conditions; some of these studies obtained rather good within-category discrimination (e.g., Aslin, Pisoni, Hennessy, & Perey, 1981; Sinnott, Beecher, Moody, & Stebbins, 1976). The spectacular success of the training studies reviewed in this subsection constitutes conclusive evidence that "...specific feedback and fixed standards in a same-different task constitute an effective procedure for the learning of acoustic cues" (Carney et al., 1977, p. 968) and that "...the utilization of acoustic differences between speech stimuli may be determined primarily by attentional factors" (p. 969).

6.1.2. Strategies and Expectations

Switching modes. We have seen that feedback and/or many hours of training are necessary to achieve a high level of within-category discrimination on a stop consonant continuum. Obviously, the acoustic differences on these continua are subtle and unfamiliar. Not only is it necessary to direct the subjects' attention to them but also subjects' discrimination accuracy needs to be sharpened by practice. There are other continua of speech sounds, however, where the acoustic differences are (or can be made) larger and more easily accessible. One might expect that little training would be necessary for acoustic discrimination of these differences, and that it would be sufficient to direct the subjects' attention to the relevant auditory dimension.

Such a case was recently investigated by Repp (1981c). He employed an /ʃ/-/s/ fricative noise continuum, followed by a vocalic context. When these stimuli were presented in AXB and fixed-standard AX tasks, most subjects perceived them fairly categorically, although within-category performance was better than expected. However, five subjects (two inexperienced and three experienced listeners) were extremely accurate in making within-category discriminations, without any specific training. Two attempts were made to teach this skill to other subjects. In one condition, the subjects were given isolated fricative noises to discriminate before listening to the fricative-vowel syllables. Although all subjects were quite accurate in detecting spectral differences in the isolated noises, their performance level dropped back to categorical levels when the noises occurred in vocalic context. In a second condition, the subjects heard a pair of noises immediately followed by exactly the same two noises in a constant vocalic context. The subjects were told to judge the isolated noises and then to verify the difference heard (if any) in the fricative-vowel syllables. Following this 25-minute training period, the subjects listened to pairs of fricative-vowel syllables only, and most subjects performed noncategorically and with high accuracy.

The success of this last procedure, together with introspections of the experienced listeners, suggested that the skill involved lay in perceptually segregating the noise from its vocalic context, which then made it possible to attend to its "pitch." Without this segregation, the phonetic percept was dominant. Once the auditory strategy has been acquired, it is possible to switch back and forth between auditory and phonetic modes of listening, and it seems likely (as Carney et al., 1977, have shown) that both strategies could be pursued simultaneously (or in very rapid succession) without any loss in accuracy. These results provide good evidence for the existence of two alternative modes of perception, phonetic and auditory--a distinction supported by much additional evidence (see Sections 5.3.3 & 5.3.4; Bailey et al., 1977; Best et al., 1981; Liberman, 1982; Repp, in press; Schwab, 1981). We may presume that the perception of other speech continua with relatively large auditory differences will likewise be susceptible to different strategies without much training.

Auditory strategies. Several studies have indicated that subjects listening to speechlike stimuli may apply different auditory strategies, given that they are operating in the auditory mode. In the phonetic mode, listeners have no choice but to integrate all the relevant acoustic information into a phonetic percept. (However, there are often individual differences in the weights given to individual cues--see, e.g., Raphael, 1981.) Once in the auditory mode, however, it is possible either to selectively attend to individual auditory dimensions or to divide attention between several of them. Thus, Best et al. (1981) found two kinds of subjects among the listeners who heard sinewave stimuli as nonspeech--"temporal listeners" and "spectral listeners" (see Section 5.3.4). However, in a recent study using speech stimuli varying along similar dimensions, Repp (1981b) found that subjects took both temporal and spectral cues into account. This divided-attention strategy was encouraged by the task that required auditory within-category discrimination (rather than auditory classification, as in Best et al., 1981).

To mention another recent example, Rosen and Howell (1981) commented on individual differences in subjects' attention to spectral and temporal cues in the discrimination of amplitude rise-time. It is not known whether there is any correlation between attentional preferences for certain cues in the auditory mode and the weights given to the same cues in phonetic perception; this seems an interesting question for future research. The availability of a variety of auditory strategies is one of the reasons why training with feedback may be required to focus subjects' attention on particular cues. However, one strategy subjects do not have available in the auditory mode is that of integrating the various cues into a single coherent percept; given that it is possible to divide attention among several cues, they remain separately perceived dimensions. Integration of psychoacoustically separable cues into a unitary percept is what characterizes the phonetic mode (Repp, 1981a, 1981b; in press). However, there are also acoustic properties that are automatically integrated in auditory perception, such as the different formants of the spectrum (Stevens & Blumstein, 1978) and that do not normally permit selective attentional strategies.

Phonetic strategies. It is also possible to adopt different strategies while operating in the phonetic mode. Such strategies take the form of shifts in the phonetic frame of reference, achieved by adding or dropping categories or even by switching to a different set altogether. Staying within the confines of a single language (see Section 6.2 for cross-linguistic research), the phonetic frame of reference for a given set of stimuli may differ from listener to listener, or it may vary within a single listener, either spontaneously or as a consequence of instructions. Of course, such variations are facilitated if the stimuli are somewhat ambiguous. There is a lot of circumstantial evidence supporting these statements, but relatively little data. However, what data there are deserve close attention because they are relevant to the question of whether or not perceptual sensitivity in a discrimination task is determined by phonetic categorization. If it is possible to shift, create, or eliminate a discrimination peak merely by applying different phonetic categories, then that peak surely cannot have a solid psychoacoustic basis.

One instructive demonstration was conducted informally by investigators at Haskins Laboratories some years ago, and although it has not found its way into the literature, it has become part of the lore. A /ba/-/da/ continuum was presented in standard identification and discrimination tasks, and the usual pronounced peak at the category boundary was obtained. Then the tests were repeated, with one minor change. That change consisted in giving the subjects the additional response category /ða/, based on the observation that synthetic syllables ambiguous between /ba/ and /da/ often sound like /ða/. (The voiced fricative /ð/ has a place of articulation intermediate between /b/ and /d/ and, in natural speech, a very weak aperiodic component that is of little perceptual significance--cf. Harris, 1958.) With the additional category (which listeners almost never use spontaneously), listeners had two category boundaries and two associated discrimination peaks, neither of which coincided with the original peak. These results provided (admittedly anecdotal) evidence for an influence of phonetic categorization per se on discrimination performance. And while it is possible to induce a similar change in categorization on a nonspeech continuum by permitting an "ambiguous" category, it is unlikely that discrimination performance will be much affected by this change (cf. Pisoni, 1977).

A recent study by Carden et al. (1981) was based on the acoustic affinity of /ba/, /da/, and /fa/, /θa/. The distinction between the two fricative categories is cued almost entirely by the vocalic formant transitions; the frication in natural productions is weak and nondistinctive (cf. Harris, 1958). Carden et al. preceded stimuli from a synthetic /ba/-/da/ continuum with a neutral noise, thus converting it into a /fa/-/θa/ continuum. The category boundaries on the two continua were significantly different. To counter the possible (though rather far-fetched) objection that the neutral noise may somehow have modified the auditory perception of the formant transitions, Carden et al. decided to hold the stimuli constant and to vary only the instructions. They first presented both continua in identification and oddity discrimination tasks, and then repeated these procedures, requiring the listeners to apply the stop categories to the fricative stimuli and vice versa. The subjects were not only able to follow these instructions, but also shifted their category boundaries in accordance with the categories used and exhibited a corresponding shift in the discrimination peak.

The results of Carden et al. provided strong evidence that the locations of the boundary and of the associated discrimination peak were not determined by psychoacoustic factors but mainly (if not exclusively) by the phonetic criteria adopted by the listeners. If there were any psychoacoustic boundaries at all on the continuum used, they seemed to be irrelevant to performance as long as the subjects operated in the phonetic mode. What seemed to matter, instead, was the relation of the stimuli to the listeners' internal "prototypes" of the relevant phonetic categories (however difficult it may be to conceptualize the mental representation of these prototypes). The difference between the /ba/-/da/ and /fa/-/θa/ boundaries is explained by the nonidentical places of articulation of these stops and fricatives, which result in characteristic differences in formant transitions. Most interestingly, it has been reported that even human infants show this boundary difference (Jusczyk, Murray, & Bayly, 1979--cited in Jusczyk, 1981). Thus, even at an early age, speech perception may not be governed solely by physical variables but may reflect an emerging (perhaps partially innate) referential system within the individual (see Section 6.3).

6.2. The Role of Linguistic Experience

Given that the degree of categorical perception in a particular experiment is largely a matter of stimulus, task, and subject factors, the central phenomenon to be explained is the phoneme boundary effect (cf. Carney et al., 1977). Cross-language research provides further valuable information on whether this effect is auditory or phonetic in origin--a question that may have no general answer and therefore must be posed separately for each particular phonetic distinction. If the effect were due to a psychoacoustic threshold, then it should not only constrain (or even pin down) the phonetic boundary locations in different languages, but it should also be associated with a discrimination peak regardless of whether or not the threshold coincides with a linguistic boundary. If the two do not coincide and perception is strongly categorical, such a peak may not be immediately evident, but it should be possible to reveal it through discrimination or classification training. On the other hand, if the phoneme boundary effect is due to phonetic categorization only, then it should occur wherever a linguistic boundary happens to be, and efforts to reveal a peak at some other fixed location should fail. It is entirely possible that phoneme boundary effects on different speech continua require different types of explanation (cf. Ades, 1977).

One obvious question one might ask is: Where are the phoneme boundaries located when subjects with different language backgrounds listen to the same continuum of synthetic stimuli? There is ample evidence from comparative phonology that category distinctions present in one language may be absent in another. Some well-known examples that will concern us below are the absence of the [ba]-[pa] (prevoiced vs. devoiced, or voiceless unaspirated) distinction in English, which is present in Thai (for example), and the absence of the /r/-/l/ distinction in Japanese (for example), which is present in English. However, there is less systematic information on the locations of boundaries between phonologically equivalent contrasts in different languages (which often differ in phonetic detail), and even less on discrimination functions corresponding to such boundaries. Since a number of relevant studies have been reviewed by Strange and Jenkins (1978), the present discussion will be brief and focus on work conducted since their article was written.

6.2.1. Cross-Linguistic Differences

By far the largest amount of cross-language work has been done on the voicing contrast for initial stop consonants, as cued by VOT. For example, Abramson and Lisker (1970; Lisker & Abramson, 1970) presented full VOT continua (containing voicing lead as well as voicing lag times) for all three places of articulation to speakers of English and Thai. The Thai subjects showed two category boundaries (prevoiced/devoiced/aspirated) and two corresponding discrimination peaks, while American listeners had only one (unaspirated/aspirated). The American and Thai results were similar on the voicing lag side (i.e., for the unaspirated-aspirated distinction common to both languages), but American listeners showed no indication of a discrimination peak on the voicing lead side, unlike Thai subjects. Similar results were obtained in a replication by Strange (1972).

Abramson and Lisker (1973) presented the same continua to speakers of Spanish, a language that distinguishes only between prevoiced and devoiced stops. The Spanish category boundaries were surprisingly close to the English ones, though at somewhat shorter voicing lag times. A major discrimination peak was obtained in the same region, together with several secondary peaks. These data contrast with a replication by Williams (1977, Fig. 1), who found the Spanish category boundary and the associated discrimination peak for labial stops to be in the vicinity of 0 msec VOT, with a secondary peak at about +25 msec of VOT, where the English /ba/-/pa/ boundary is located. While the discrepancy between these two studies remains unexplained, Williams' results--which appear more reliable--are interesting for two reasons: First, they show that Spanish listeners can accurately discriminate between VOT values in the very short lead/lag range where, according to psychophysical arguments (Pisoni, 1977), they should be limited to near-chance performance by the simultaneity-successiveness threshold. Second, the secondary peak at short lag times suggests that these listeners were able to discriminate unaspirated from aspirated stops, presumably on an auditory basis. If so, then discrimination at very short VOTs was either entirely phonetic in nature (i.e., based on subjective uncertainty of phonetic judgments) or based on spectral signal properties (cf. Samuel, 1977), while the secondary peak at short lag times may have represented the temporal-order threshold postulated by Pisoni (1977). The ability of Spanish listeners to discriminate unaspirated from aspirated stops contrasts with English-speaking listeners' inability to spontaneously discriminate prevoiced from devoiced stops. Presumably, the presence of prevoicing is less salient at the psychoacoustic level than the presence of aspiration (with its higher amplitude and concomitant spectral changes in the signal).

In a recent study of Polish, whose stop categories resemble those of Spanish, Keating, Mikos, and Ganong (1981) found a VOT boundary in the short lag range (close to zero VOT), together with a very broad discrimination peak that was skewed towards longer lag times. They also found that the boundary could be shifted towards longer voicing lags by adjusting the stimulus range so it included more aspirated tokens. These results suggest, in accord with the Spanish findings, that the presence of aspiration is a rather salient auditory event. Williams (1977) also found a broad discrimination peak similar to the Polish one for several Spanish-English bilinguals.

One phenomenon that has attracted the attention of researchers for some time is the inability of Japanese subjects to distinguish (and to correctly produce) American English /r/ and /l/, neither of which occurs in Japanese. (The Japanese /r/ is a dental flap--see Price, 1981.) These difficulties often persist for individuals who are quite fluent in English (Goto, 1971). An experimental demonstration was provided by Miyawaki et al. (1975), who showed that Japanese subjects performed very poorly when labeling or discriminating stimuli from a synthetic /ra/-/la/ continuum that were perceived fairly categorically by American listeners. However, when the distinctive third formants of these stimuli were presented in isolation as a nonspeech control, Japanese and American listeners gave almost identical results, with discrimination performance clearly above chance. This result suggested that the effect of linguistic experience was restricted to perception in the speech mode.

Little direct cross-language research has been done on other phonetic contrasts. For example, virtually nothing is known about the effect of linguistic background on the perception of stop consonant place of articulation. Stevens et al. (1969) compared American and Swedish listeners' perception of steady-state vowels. Although there were differences in the locations of category boundaries, they were not reflected in the discrimination functions, which were very similar for the two groups of listeners. This study is well worth repeating, in view of consistent findings of discrimination peaks at vowel category boundaries. Thus, for example, the Japanese subjects of Fujisaki and Kawashima (1969, 1970) show a single discrimination peak on an /i/-/e/ continuum, while American listeners show two peaks on a very similar continuum (Pisoni, 1971: Exp. 1), on which they distinguish three categories (/i/, /I/, /ɛ/).

A cross-language difference in fricative perception may be gleaned from a comparison of data by Kunisaki and Fujisaki (1977) for Japanese listeners, and by Repp (1981c) for American listeners. Both studies used rather similar /ʃ/-/s/ continua, but the locations of the Japanese and American boundaries are different, and both are associated with marked discrimination peaks (cf. Fujisaki & Kawashima, 1969). Other comparisons of this sort, between separate studies conducted in different countries, could probably be found.

6.2.2. Acquisition of a New Phonetic Contrast

Students of a foreign language encounter the problem of learning to perceive and produce unfamiliar phonetic contrasts. Considering the importance of this problem, it is surprising how little laboratory research it has generated. The few studies in the literature were again concerned with either VOT or the /r/-/l/ contrast.

Given listeners' apparent sensitivity to the presence of aspiration in syllable-initial stops, it should be easy to teach Spanish or Polish listeners to discover the unaspirated-aspirated distinction. Lisker (1970) trained Russian listeners to discriminate labial stops ranging in VOT from +10 to +60 msec, all of which they normally label "p." The subjects learned to attach different labels to the endpoints of this range, but when labeling the stimuli in between, they showed a rather gradual change with a mid-range boundary that did not correspond to the American boundary (which is at about 25 msec). No discrimination tests were administered. Lisker concluded that Russian and American listeners used different criteria for judging the same stimuli, with the Russians exhibiting either continuous perception or a different "natural" boundary in the voicing lag region. Pisoni et al. (1982) later criticized Lisker's study for not having employed feedback, thereby perhaps not directing the subjects' attention to the "correct" acoustic cues. They cite a study by Lane and Moore (1962), who successfully employed training with feedback to teach an aphasic patient the re-acquisition of the English voicing contrast, using the /do/-/to/ (F1 cutback) continuum of Liberman, Harris, Kinney, and Lane (1961). Unfortunately, there have been no further studies with Russian subjects.

Several studies have attempted to teach American listeners the prevoiced-devoiced distinction for which they show little spontaneous sensitivity. After having relatively little success with extensive training in oddity discrimination, Strange (1972) first taught listeners to associate arbitrary labels with a clearly prevoiced (-100 msec VOT) and a clearly devoiced (+10 msec VOT) stop before administering standard identification and oddity discrimination tests, using the negative VOT range only. The subjects showed fairly orderly labeling functions and improved discrimination scores following training, but the location of the category boundary was variable, and so were the shapes of the discrimination functions. Moreover, there was no transfer of training from an alveolar to a labial VOT continuum. Comparably variable results were obtained in a second study that provided training in judging VOT stimuli on a continuous scale.

Pisoni et al. (1982) resumed the task abandoned by Strange, with quite different results. They quite simply asked naive subjects to use "three response categories corresponding to [b], [p] and [p^h]" (p. 301) and obtained surprisingly consistent labeling in the prevoicing region, even without any special training (although training improved labeling consistency). What may have been responsible for their success but, curiously, was not mentioned by Pisoni et al. (but see McClasky, Pisoni, & Carrell, 1980), was that the categories used by the subjects were in fact "mba," "ba," and "pa." Apparently, it helped a great deal to associate the unfamiliar prevoicing distinction with a familiar phonemic contrast (even though initial nasal-stop clusters do not occur in English). In ABX discrimination tests, two peaks were found--a major one at the regular category boundary at short voicing lags (+20 msec of VOT), and a minor one in the short voicing lead region (-20 msec of VOT). Interestingly, both peaks were obtained regardless of whether or not the subjects had any prior labeling experience, either with two or with three categories. This finding contrasts with previous data that had found no discrimination peak in the voicing lead region. One factor that may have played a role here is the amplitude of the prevoicing, which may have been higher in the Pisoni et al. stimuli. (No amplitudes are mentioned in any of the studies.) There is no doubt that the detectability and discriminability of prevoicing will increase with its amplitude.

It is by no means clear that the new category distinction acquired by the subjects of Pisoni et al. (1982), even though it was apparently precipitated by the use of phonetic labels, was indeed a phonetic one (or, if it was, that it was the prevoiced/devoiced rather than the nasal+stop/stop distinction). The "mba" label may simply have served to direct the subjects' attention to the relevant auditory dimension. A subsequent demonstration by McClasky et al. (1980) of virtually perfect transfer of the acquired distinction to an alveolar stop ("nda"-"da") continuum proves little, for the prevoiced portion is acoustically independent of the place of articulation of the stop consonant. The critical question is whether subjects who are able to perceive the prevoicing distinction in the laboratory will subsequently be able to use this skill in a natural-language context, e.g., in learning a foreign language like Thai. Until such transfer has been demonstrated, it is prudent to assume that the subjects of Pisoni et al., rather than acquiring a new phonetic contrast, merely learned to make certain auditory discriminations.

The importance of conducting discrimination training in a way that facilitates transfer to a more naturalistic situation was stressed by MacKain et al. (in press), who re-examined Japanese listeners' perception of the English /r/-/l/ distinction. They found several individuals who were able to identify and discriminate stimuli from a /rɑk/-/lɑk/ ("rock"-"lock") continuum almost as well (i.e., as categorically) as American subjects. It turned out that these subjects had not only had extensive experience with English but with English conversation in particular, suggesting that transfer from the real world to the laboratory may be easier than the other way around. The continuing research in this area promises to yield useful insights into the process of second-language acquisition.

6.3. Categorical Perception in Human Infants

Since the rather extensive literature on infant speech perception has been reviewed repeatedly in recent years (Eilers, 1980; Juscyzk, 1981, in press; Kuhl, 1979b; Mehler & Bertoncini, 1979; Morse, 1979; Walley, Pisoni, & Aslin, 1981), only a very brief summary is needed here. It is now well known that infants as young as a few weeks do exhibit categorical discrimination. Although, for obvious methodological reasons, this result is usually established with a much smaller number of different stimuli than are used in corresponding studies with adult subjects, the pattern is generally clear: Pairs of stimuli crossing the adult (American English) boundary are discriminated more readily than pairs of stimuli from within an adult category. This has been shown for the voicing lag (unaspirated-aspirated) contrast in initial stop consonants (Eimas et al., 1971; however, see Molfese & Molfese, 1979), for the place-of-articulation contrast in voiced initial stop consonants (Eimas, 1974), for the /ra/-/la/ distinction (Eimas, 1975), and for the /ba/-/wa/ distinction (Eimas & Miller, 1980). Isolated vowels, on the other hand, appear to be continuously discriminated by infants (Swoboda, Kass, Morse, & Leavitt, 1978).

In addition, there are a number of studies that, while not testing for within-category discrimination, have demonstrated the infant's ability to discriminate a variety of phonetic contrasts in natural or synthetic speech (e.g., Jusczyk, 1977; Jusczyk, Copan, & Thompson, 1978; Jusczyk & Thompson, 1978). Categorical-like discrimination has also been found for Pisoni's (1977) tone-onset-time continuum (Jusczyk, Pisoni, Walley, & Murray, 1980), while isolated third formants from a /ra/-/la/ continuum (Miyawaki et al., 1975) were perceived continuously by infants (Eimas, 1975). With the exception of occasional negative findings due to procedural factors (see Morse, 1979) or to the difficulty of certain phonetic contrasts (e.g., /f/-/θ/, Eilers, Wilson, & Moore, 1977), these results show the infant's perceptual capabilities to be remarkably developed and broadly similar to those of adults.

One important difference, however, is that infants have only minimal linguistic experience. It is generally considered unlikely that a few weeks or months of passive exposure to a particular language could have any significant effect on the infant's perceptual response to speech stimuli. Thus, infants reared in different language environments are expected to behave similarly, and this expectation has been confirmed in several cross-linguistic studies. What makes these studies especially interesting is that they show infants to be sensitive to certain distinctions that are not phonemic in their future language. Thus, American infants apparently can discriminate the prevoiced-devoiced contrast (Aslin et al., 1981; Eimas, 1975), while Kikuyu (Streeter, 1976) and Spanish infants (Eilers, Gavin, & Wilson, 1979; Lasky, Syrdal-Lasky, & Klein, 1975) can discriminate the unaspirated-aspirated contrast, which does not figure in their respective languages. While it has not been established that infants perceive these "unfamiliar" distinctions in a truly categorical fashion (cf. Aslin et al., 1981; Morse, 1979), these results, at the very least, demonstrate high sensitivity to certain auditory stimulus properties--a sensitivity that adults seem to suppress unless these properties become associated with a phonetic distinction.

Additional evidence for American infants' superiority over adults in discriminating foreign-language contrasts has been obtained by Trehub (1976) for vowel nasalization and fricative palatalization, by Werker, Gilbert, Humphrey, and Tees (1981) for the dental-retroflex and aspirated voiced-voiceless contrasts, and by Werker (1982) for the dental-retroflex and velar-uvular contrasts. The work of Werker (1982) is especially intriguing in that it has provided longitudinal evidence that the ability to discriminate these contrasts disappears as early as 8-10 months of age, a time at which recognizable phonetic segments emerge in babbling. This startling finding has recently been confirmed in a longitudinal study of individual infants (Werker, personal communication).

Of course, these findings should not be interpreted as showing that infants' auditory sensitivity is superior to that of adults. In fact, the opposite is likely to be the case; for example, higher tone-onset-time thresholds have been obtained with infants than with adults (Jusczyk et al., 1980) and, in a recent comparison of VOT discrimination thresholds obtained with identical procedures (Aslin et al., 1981), adults proved to be far superior to infants. However, infants are free to attend to auditory properties of speech while adults, being constrained by linguistic experience, are not. Once adults' attention is properly directed to auditory stimulus attributes (see Section 6.1.2), their discrimination performance is likely to be superior to that of infants.

The infant research has also revealed instances of phonetic distinctions that are not discriminated at an early age but are contrastive in the language. One such distinction is that between short negative and short positive VOTs, which crosses a phoneme boundary in Spanish but not in English (Lasky et al., 1975). Presumably, infants in a Spanish-speaking environment must learn this distinction as they grow older, while learning to disregard other distinctions that are not phonemic in their language. Thus, this research again attests to the profound influence that linguistic experience exerts on speech perception. What is not yet clear is whether the infant's perceptual predispositions are purely auditory in nature, or whether they already reflect specifically linguistic propensities. Recent research on trading relations between different acoustic speech cues in infants suggests the possibility of some innate linguistic mechanisms (Miller & Eimas, in press), as does the finding of different boundaries on /ba/-/da/ and /fa/-/θa/ continua (Jusczyk et al., 1979, cited in Jusczyk, 1981). Just how specific these mechanisms are and how they interact with later experience remains to be investigated in more detail. For excellent discussions of issues in the development of speech perception, see Aslin and Pisoni (1980) and Jusczyk (in press).

6.4. Categorical Perception in Nonhuman Animals

The question of whether human infants are endowed with any specific genetic predispositions for phonetic perception is usefully addressed by comparing their speech perception with that of nonhuman animals. Unless an animal has had extensive experience with human speech (and probably even then), its ability to discriminate speech sounds should reflect solely psychoacoustic factors. Provided that its auditory system is similar to the human one (which is true for the two species studied most closely, macaques and chinchillas), the results from the animal laboratory should reveal how much of the human infant's performance can be attributed to pure psychoacoustics.

Because of obvious methodological difficulties, animal research on speech perception has made only slow progress. A recent article (Kuhl, 1981) cites only four earlier studies concerned with categorical perception.

Morse and Snowdon (1975) measured changes in macaques' heart rate in response to changes in speech stimuli drawn from Pisoni's (1971: Exp. I) /bæ/-/dæ/-/gæ/ continuum. The monkeys exhibited good discrimination between categories, and also some sensitivity to within-category differences, although the latter finding rested primarily on an unexplained heart-rate acceleration in the no-change control condition. Sinnott et al. (1976) tested macaques and humans on a /ba/-/da/ continuum, using a key-press response and a fixed-standard paradigm. While the results for humans were not very categorical (humans were actually better than monkeys in detecting within-category differences), those for the monkeys did not suggest categorical perception either. Because of differences in procedure, these results are not easily compared with those of Morse and Snowdon. Waters and Wilson (1976) used avoidance training to test macaques' discrimination of stimuli from a VOT continuum. Their data, like those of Sinnott et al., yielded only the equivalent of labeling functions obtained with several different ranges of VOT. The monkeys' "category boundary" was found to be highly range-dependent, which suggests continuous perception. Since the boundary was consistently located in the voicing lag region, it seems likely that the animals paid attention to the presence of aspiration noise or to spectral differences in the F1 region.

Of these three studies, only that by Morse and Snowdon (1975) provides some indication of a category boundary effect in monkeys. Clearly, those data need to be replicated if they are to stand on solid ground. However, a highly successful demonstration of category boundary effects in monkeys has recently been reported (Kuhl & Padden, 1982a).

Animals would be expected to show categorical perception of speech only when a speech continuum straddles a psychoacoustic threshold. This may be true for the VOT continuum. In a widely cited study, Kuhl and Miller (1978) reported almost identical "labeling functions" (i.e., generalization gradients) for chinchillas and for humans on three VOT continua, /ba/-/pa/, /da/-/ta/, and /ga/-/ka/. For both groups of subjects, the boundaries shifted towards longer values of VOT as place of articulation changed from labial to alveolar to velar, even though the range of VOTs remained constant. These results strongly suggested a psychoacoustic reason for the boundary shift, probably due to the spectral concomitants of VOT. No attempt was made to test whether the chinchilla boundary is as stable with changes in stimulus range as the human boundary (cf. Brady & Darwin, 1978; Keating et al., 1981) or as unstable as the monkey boundary (Waters & Wilson, 1976).

Discrimination data for chinchillas were recently reported by Kuhl (1981). After training the animals to avoid shock by responding to differences between successive stimuli, she used a staircase procedure to determine VOT difference limens at various points along a /da/-/ta/ continuum. She found the highest accuracy in the region between 30-40 msec of VOT, where both the human and the chinchilla boundaries are also located. A previous unpublished study by Miller, Henderson, Sullivan, and Rigden (1978) had shown superior discrimination of stimuli crossing the boundary on a /ga/-/ka/ continuum. These results provide rather strong evidence of a psychoacoustic boundary in the voicing lag region for chinchillas (and, presumably, for humans as well). Similar results have recently been obtained with monkeys (Kuhl & Padden, 1982b). What remains uncertain is the role of these psychoacoustic factors in human speech perception. We agree with Pisoni's (1980b) reservation that findings on animal speech perception "...are incapable, in principle, of providing any further information about how these signals might be 'interpreted' or coded within the context of the experience and history of the organism" (p. 304).

7. CONCLUDING COMMENTS: BEYOND THE CATEGORICAL PERCEPTION PARADIGM

The research reviewed in the preceding sections has operated almost exclusively within a single experimental paradigm. Although there have been a great many variations in procedural detail, the essential common factor has been the use of (typically synthetic) continua of speech sounds. This concluding section offers some comments on the limitations of this approach, and on its relation to categorical perception in the real world.

7.1. On Articulatory Realism

The possibility of constructing a continuum from one phonetic category to another is intriguing. However, the stimuli on such a continuum are not all equally realistic. While the endpoint stimuli of a synthetic continuum are already removed from real speech by virtue of their stylized acoustic properties, this is even more true for stimuli from the middle of the continuum, which were never intended to model real speech but were obtained by mere parameter interpolation. In some cases, utterances resembling these stimuli may actually be impossible to produce by a human vocal tract.

While this argument may be used to downgrade categorical perception research for its lack of ecological realism, it has not been traditionally considered a disadvantage. Indeed, it is part and parcel of the "motor-theoretic" view of categorical perception: Perception is categorical where the articulatory space (in a given language) is relatively discontinuous--in other words, when the stimuli from the middle of a continuum are less realistic than those from the ends. Seen in this way, the motor theory is not so much a theory as a statement of (though often poorly documented) fact. The mechanisms by which perceptual processes might "refer to" articulation have always remained obscure, which has led many researchers to dismiss the motor theory altogether. Nobody would deny, however, that perception is shaped by experience, and that this shaping is due to events that occur frequently. Therefore, the phonetic categories that constitute the frame of reference for speech perception _must_ directly reflect the structure of speech--a structure that is imposed by the articulatory system within the conventions specific to a given language. Consequently, it is a truism that speech perception is intimately related to speech production. How this relationship is instantiated and solidified in the brain is a question for the philosopher and the neurophysiologist to answer. (For some interesting developments in the latter direction, see Anderson, Silverstein, Ritz, & Jones, 1977.) The difficulty of finding an answer should not prevent us, however, from recognizing that the specific systemic properties of speech are equally reflected in production and perception.

Several theorists have argued that, when listening to speech, we directly perceive what the articulators are doing (e.g., Gibson, 1966; Neisser, 1976; Summerfield, 1979). Essentially, this hypothesis is a contemporary version of the motor theory, though it denies any role of "mediation" or "reference" in perception. As far as natural speech is concerned, the hypothesis must be true, for speech _is_ what the articulators are doing, as conveyed by sound. However, this cannot be said of the stimuli from synthetic continua. To the extent that they are unlikely products of articulation, they should be perceived either as nonspeech or be perceptually assimilated to existing schemata of articulatory action, which are instantiated by the phonetic categories of a language. The phenomenon of categorical perception suggests that, as long as the stimuli capture some salient properties of speech, they are perceived as the articulatory event most compatible with their structure, and this seems consistent with theories of direct perception, particularly with Neisser's (1976) formulation.

7.2. _On Category Boundaries_

The view of categorical perception as an acquired, language-specific, attentional phenomenon seems to contradict the hypothesis that categorical perception is caused by psychophysical boundaries on a stimulus continuum. However, the contradiction is more apparent than real. There is extensive evidence, reviewed above, that categorical perception may be caused _either_ by categorization alone _or_ by a psychophysical discontinuity, and that both factors may be operating simultaneously for a single set of stimuli (although the former seems much more important in speech perception than the latter). Problems arise only when it is attempted to reduce these two causes to a single one, by assuming that auditory thresholds are plastic and shift with language experience (see, e.g., Aslin & Pisoni, 1980). This hypothesis (which is forced by the common-factor theory of categorical perception) is empty if the auditory thresholds in question are assumed to be entirely specific to speech, i.e., if they are essentially equated with phonetic boundaries; and it is most likely wrong if auditory thresholds are understood in a more general sense. In the second case, for example, the thresholds for certain nonspeech distinctions should show language-specific variations along with the phonetic boundaries they are presumed to underlie--a prediction for which there is currently no positive evidence whatsoever. It seems much more likely that auditory thresholds and phonetic boundaries coexist, with the former limiting the possible locations of the latter only in the sense that what sounds the same cannot be phonetically distinctive.

One true shortcoming of the categorical perception paradigm is that it has overemphasized the importance of the boundaries between phonetic categories. After all, the categories, and not the boundaries between them, are the important functional elements of speech and language. The boundaries themselves are a mere epiphenomenon, apparent only in a particular experimental situation. Within the limits of the categorical perception paradigm, it may often not be clear whether the boundary is there because of the categories or whether the categories are there because of the boundary (although it should be possible, at least in principle, to decide this issue empirically in each case). However, beyond the realm of artificial speech continua, the boundary concept has little to offer.

It is appropriate to mention at this point some interesting research concerned with the basis of linguistic categories per se, disregarding the question of boundaries. For example, Fodor, Garrett, and Brill (1975) reinforced infants to respond with head turns to two (out of three) CV syllables that either did or did not share the initial consonant, the vowels always being different. The infants showed more evidence of learning when the consonants were shared, indicating some ability to detect invariant acoustic properties (cf. Stevens & Blumstein, 1978) or perhaps even to conduct some sort of segmental analysis (Fodor et al., 1975). Kuhl (1979a) demonstrated that infants are able to respond differentially to two vowel categories (/a/ and /i/) in the presence of a wide variety of distracting variability (different talkers). Similar perceptual constancy for vowels, at least, has been demonstrated in dogs (Baru, 1975) and chinchillas (Burdick & Miller, 1975). Perceptual classification techniques of this kind have also been used with adults to examine the possible psychoacoustic basis for the perceived similarity of stop consonants in initial and final position (Grunke & Pisoni, 1979; Schwab, 1981) or across different vocalic contexts (Jusczyk, Smith, & Murphy, 1981), as well as listeners' awareness of phonological features (Healy & Levitt, 1980). These and related methods promise to provide useful information, particularly about the emergence of phonetic categories in human infants, without undue emphasis on the boundaries between categories.

7.3. On Dual Processing

Several recent reviews have argued that the dual-process hypothesis of categorical perception should be abandoned in favor of single-process models (e.g., Crowder, 1981; Macmillan et al., 1977; Tartter, 1982). While it is true that the results of particular experiments are sometimes difficult to decompose into separate contributions of phonetic and auditory judgments, the basic distinction between the two modes of processing is logically unassailable (Pisoni, 1980b; Repp, in press). To classify stimuli into the categories characteristic of the language is simply different from judging stimuli as long or short, constant or changing, continuous or interrupted, etc. We have reviewed several experiments showing that listeners can switch between phonetic and auditory modes, with often strikingly different results. There is no reason to doubt the original suggestion of Fujisaki and Kawashima (1969, 1970) that both modes may be employed simultaneously in a discrimination task; whether they are, depends on the specific situation.

Categorical perception of speech is, first and foremost, an experimental demonstration that listeners persist in their normal perceptual habits in the laboratory, even when given the opportunity to relinquish those habits. There is nothing surprising about the categorical nature of speech perception, which was known long before the discovery of the laboratory phenomenon of categorical perception. The interest of the phenomenon lies solely in subjects' strong resistance to adopt a mode of listening that enables them to detect subphonemic detail. That this resistance can be overcome by appropriate methods and training is one of the most significant findings reviewed here. An important question for future research will be whether analytic perceptual skills acquired in the laboratory can be transferred to real-life situations. However, the question immediately comes to mind: Having trained subjects to overcome their language habits and to pay some attention to the sound of speech, of what use could that esoteric skill be to them in the real world?

There are two (related) real-life endeavors that require the (more or less conscious) apprehension of subphonemic distinctions. One is phonetic transcription; the other is acquisition of a foreign language. Phonetic transcription is a skill that phoneticians acquire through training. However, even in its more narrow varieties, it is essentially categorization according to a fine-grained scheme, instantiated by the International Phonetic Alphabet. Thus, rather than paying attention to auditory properties of speech, phoneticians simply use a larger number of internalized phonetic categories than the ordinary individual. However, phoneticians are usually also able to make some fairly accurate judgments about the auditory quality of speech sounds. That such an ability could be cultivated to a high degree is presupposed in Pilch's (1979) proposal of a science of "auditory phonetics," which involves the systematic description, using a purely auditory vocabulary, of "the partitions of auditory space imposed by different phonemic systems" (p. 157). While, for purposes of communication, the auditory description once again makes use of categories, these categories are intended to be decidedly nonphonetic. How succcessful this approach will be, given the twin difficulties of attending to auditory properties of speech in a natural setting and of finding the proper terms for their description, remains to be seen. It is possible, however, that laboratory training of the sort employed in several recent categorical-perception studies (e.g., Carney et al., 1977) will be helpful in developing the auditory phonetician's skills. Such skills may also be useful to speech pathologists.

A similar (and more commonly encountered) problem faces the individual learning a foreign language. In order to detect certain novel phonetic distinctions and to realize them in production, some sensitivity to subphonemic detail is required (cf. Flege, 1981; Flege & Hammond, in press). Note, however, that at no time does the language learner need to describe this detail in auditory terms, or to detect differences that are subphonemic in both the new and old languages. The task is restricted to the acquisition of new phonetic categories--a process that may not involve the auditory mode of perception at all, at least not at the level of consciousness. The possibility that an increased awareness of the auditory properties of speech might facilitate the acquisition of new phonetic contrasts outside the laboratory certainly deserves continued attention, but we should perhaps not be overly optimistic. So far, there is no convincing evidence that new phonetic contrasts can be taught directly in the laboratory by the simple techniques discussed here. A fruitful connection between categorical perception research and foreign language instruction still needs to be made.

The prospect of gaining some insight into the processes of both first- and second-language acquisition will keep interest in the phenomenon of categorical perception alive. It is to be expected, however, that the traditional methodology will eventually give way to new approaches that more directly address the important theoretical and practical problems raised by communication in the real world. Indeed, it seems that this process is now well under way.

REFERENCES

Abramson, A. S. Identification and discrimination of phonemic tones. *Journal of the Acoustical Society of America*, 1961, *33*, 842. (Abstract)

Abramson, A. S. The noncategorical perception of tone categories in Thai. In B. Lindblom & S. Öhman (Eds.), *Frontiers of speech communication research*. London: Academic Press, 1979, pp. 127-134.

Abramson, A. S., & Lisker, L. Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the Sixth International Congress of Phonetic Sciences*. Prague: Academia, 1970, pp. 569-573.

Abramson, A. S., & Lisker, L. Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics*, 1973, *1*, 1-8.

Abramson, A. S., Nye, P. W., Henderson, J. B., & Marshall, C. W. Vowel height and the perception of consonantal nasality. *Journal of the Acoustical Society of America*, 1981, *70*, 329-339.

Ades, A. E. *Categorical perception and the speech mode*. Unpublished manuscript, MIT, 1973.

Ades, A. E. Adapting the property detectors for speech perception. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms*. Amsterdam: North Holland, 1976.

Ades, A. E. Vowels, consonants, speech, and nonspeech. *Psychological Review*, 1977, *84*, 524-530.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, 1977, *84*, 413-451.

Aslin, R. N., & Pisoni, D. B. Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. 2). New York: Academic Press, 1980, pp. 67-96.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L., & Perey, A. J. Discrimination of voice onset time by human infants: New findings and implications for the effect of early experience. *Child Development*, 1981, *52*, 1135-1145.

Bailey, P. J., & Summerfield, Q. Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, *6*, 536-563.

Bailey, P. J., Summerfield, Q., & Dorman, M. On the identification of sine-wave analogues of certain speech sounds. *Haskins Laboratories Status Report on Speech Research*, 1977, *SR-51/52*, 1-25.

Barclay, J. R. Noncategorical perception of a voiced stop: A replication. *Perception & Psychophysics*, 1972, *11*, 269-273.

Baru, A. V. Discrimination of synthesized vowels [a] and [i] with varying parameters (fundamental frequency, intensity, duration, and number of formants) in dog. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech*. New York: Academic Press, 1975, pp. 91-101.

Bastian, J., & Abramson, A. S. Identification and discrimination of phonemic vowel duration. *Speech Research and Instrumentation* (Tenth Final Report). New York: Haskins Laboratories, 1964. (Appendix L)

Bastian, J., Eimas, P., & Liberman, A. M. Identification and discrimination of a phonemic contrast induced by silent interval. *Journal of the Acoustical Society of America*, 1961, *33*, 842. (Abstract)

Baumrin, J. M. Perception of the duration of a silent interval in nonspeech stimuli: A test of the Motor Theory of speech perception. *Journal of Speech and Hearing Research*, 1974, *17*, 294-309.

Berlin, B., & Kay, P. *Basic color terms: Their universality and evolution*. Berkeley: University of California Press, 1969.

Best, C. T., Morrongiello, B., & Robson, R. Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, 1981, *29*, 191-211.

Blechner, M. J. *Musical skill and the categorical perception of harmonic mode*. Unpublished doctoral dissertation, Yale University, 1977.

Blumstein, S. E., & Stevens, K. N. Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America*, 1980, *67*, 648-662.

Bornstein, M. H. Color vision and color naming: A psychophysiological hypothesis of cultural difference. *Psychological Bulletin*, 1973, *80*, 257-285.

Bornstein, M. H., Kessen, W., & Weiskopf, S. Color vision and hue categorization in young human infants. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, *2*, 115-129.

Brady, P. T., House, A. S., & Stevens, K. N. Perception of sounds characterized by a rapidly changing resonant frequency. *Journal of the Acoustical Society of America*, 1961, *33*, 1357-1362.

Brady, S., & Darwin, C. J. Range effect in the perception of voicing. *Journal of the Acoustical Society of America*, 1978, *63*, 1556-1558.

Braida, L. D., & Durlach, N. I. Intensity perception: II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America*, 1972, *51*, 483-502.

Burdick, C. K., & Miller, J. D. Speech perception by the chinchilla: Discrimination of sustained /a/ and /i/. *Journal of the Acoustical Society of America*, 1975, *58*, 415-427.

Burns, E. M., & Ward, W. D. Categorical perception -- phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, 1978, *63*, 456-468.

Carden, G., Levitt, A., Jusczyk, P. W., & Walley, A. Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place. *Perception & Psychophysics*, 1981, *29*, 26-36.

Carney, A. E., Widin, G. P., & Viemeister, N. F. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 1977, *62*, 961-970.

Carrell, T. D., Pisoni, D. B., & Gans, S. J. Perception of the duration of rapid spectrum changes: Evidence for context effects with speech and nonspeech. *Research on Speech Perception* (Progress Report No. 6). Bloomington, IN: Indiana University, Department of Psychology, 1980, pp. 421-436.

Chistovich, L. A. Problems of speech perception. In L. L. Hammerich, R. Jakobson, & E. Zwirner (Eds.), *Form & substance*. Copenhagen: Akademisk Forlag, 1971, pp. 83-93.

Chistovich, L. A., Fant, G., de Serpa-Leitão, A., & Tjernlund, P. Mimicking and perception of synthetic vowels. *Quarterly Progress and Status Report*. Stockholm: Royal Technical University, Speech Transmission Laboratory, 1966, *2*, 1-18.

Conway, D. A., & Haggard, M. P. New demonstrations of categorical perception. *Speech Synthesis and Perception* (Report No. 5). Cambridge, England: Psychology Laboratory, 1971, pp. 51-73.

Cooper, W. E. Selective adaptation of speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory*. Hillsdale, N.J.: Erlbaum, 1975.

Cooper, W. E., Ebert, R. E., & Cole, R. A. Perceptual analysis of stop consonants and glides. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, *2*, 92-104.

Cowan, N., & Morse, P. A. Influence of task demands on the categorical versus continuous perception of vowels. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers*. New York: Acoustical Society of America, 1979, pp. 443-446.

Creelman, C. D., & Macmillan, N. A. Auditory phase and frequency discrimination: A comparison of nine procedures. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 146-156.

Cross, D. V., Lane, H. L., & Sheppard, W. C. Identification and discrimination functions for a visual continuum and their relation to the motor theory of speech perception. *Journal of Experimental Psychology*, 1965, *70*, 63-74.

Crowder, R. G. Improved recall for digits with delayed recall cues. *Journal of Experimental Psychology*, 1969, *82*, 258-262.

Crowder, R. G. The sound of vowels and consonants in immediate memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 587-596.

Crowder, R. G. Precategorical acoustic storage for vowels of short and long duration. *Perception & Psychophysics*, 1973, *13*, 502-506. (a)

Crowder, R. G. Representation of speech sounds in precategorical acoustic storage. *Journal of Experimental Psychology*, 1973, *98*, 14-24. (b)

Crowder, R. G. Mechanisms of auditory backward masking in the stimulus suffix effect. *Psychological Review*, 1978, *85*, 502-524.

Crowder, R. G. The role of auditory memory in speech perception and discrimination. In T. Myers, J. Laver, & J. Anderson (Eds.), *The cognitive representation of speech*. Amsterdam: North-Holland, 1981, pp. 167-180.

Crowder, R. G. Decay of auditory memory in vowel discrimination. *Journal of Experimental Psychology: Human Learning and Memory*, 1982, *8*, 153-162. (a)

Crowder, R. G. The communality of auditory sensory storage in perception and immediate memory. *Perception & Psychophysics*, 1982, *31*, 477-483. (b)

Crowder, R. G., & Morton, J. Pre-categorical acoustic storage (PAS). *Perception & Psychophysics*, 1969, *5*, 365-373.

Cutting, J. E. Two left-hemisphere mechanisms in speech perception. *Perception & Psychophysics*, 1974, *16*, 601-612.

Cutting, J. E. There may be nothing peculiar to perceiving in a speech mode. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J.: Erlbaum, 1978, pp. 229-244.

Cutting, J. E., & Rosner, B. S. Categories and boundaries in speech and music. *Perception & Psychophysics*, 1974, *16*, 564-570.

Cutting, J. E., & Rosner, B. S. Discrimination functions predicted from categories in speech and music. *Perception & Psychophysics*, 1976, *20*, 87-88.

Cutting, J. E., Rosner, B. S., & Foard, C. F. Perceptual categories for musiclike sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology*, 1976, *28*, 361-378.

Darwin, C. J., & Baddeley, A. D. Acoustic memory and the perception of speech. *Cognitive Psychology*, 1974, *6*, 41-60.

Dechovitz, D., & Mandler, R. Effects of transition length on identification and discrimination along a place continuum. *Haskins Laboratories Status Report on Speech Research*, 1977, *SR-51/52*, 119-130.

Diehl, R. L. Feature detectors for speech: A critical reappraisal. *Psychological Bulletin*, 1981, *89*, 1-18.

Diehl, R. L., Elman, J. L., & McCusker, S. B. Contrast effects on stop consonant identification. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, *4*, 599-609.

Dorman, M. F. Auditory evoked potential correlates of speech sound discrimination. *Perception & Psychophysics*, 1974, *15*, 215-220.

Dorman, M. F., & Dougherty, K. Shifts in phonetic identification with changes in signal presentation level. *Journal of the Acoustical Society of America*, 1981, *69*, 1439-1440.

Durlach, N. I., & Braida, L. D. Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, 1969, *46*, 372-383.

Edman, T. R. Discrimination of intraphonemic differences along two place of articulation continua. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers*. New York: Acoustical Society of America, 1979, pp. 455-458.

Edman, T. R., Soli, S. D., & Widin, G. P. Learning and generalization of intraphonemic VOT discrimination. *Journal of the Acoustical Society of America*, 1978, *63* (Suppl. No. 1), S19. (Abstract)

Eilers, R. E. Infant speech perception: History and mystery. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. II). New York: Academic Press, 1980, pp. 23-39.

Eilers, R. E., Gavin W., & Wilson, W. R. Linguistic experience and phonemic perception in infancy: A cross-linguistic study. *Child Development*, 1979, *50*, 14-18.

Eilers, R. E., Wilson, W. R., & Moore, J. M. Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research*, 1977, *20*, 766-780.

Eimas, P. D. The relation between identification and discrimination along speech and non-speech continua. *Language and Speech*, 1963, *6*, 206-217.

Eimas, P. D. Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics*, 1974, *16*, 513-521.

Eimas, P. D. Auditory and phonetic coding of the cues for speech: Discrimination of the [r-l] distinction by young infants. *Perception & Psychophysics*, 1975, *18*, 341-347.

Eimas, P. D., & Corbit, J. D. Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 1973, *4*, 99-109.

Eimas, P. D., & Miller, J. L. Auditory memory and the processing of speech. *Developmental Studies of Speech Perception* (Progress Report No. 3). Providence, RI: Brown University, 1975, pp. 117-135.

Eimas, P. D., & Miller, J. L. Effects of selective adaptation on the perception of speech and visual patterns: Evidence for feature detectors. In H. L. Pick & R. D. Walk (Eds.), *Perception and experience*. New York: Plenum Press, 1978.

Eimas, P. D., & Miller, J. L. Contextual effects in infant speech perception. *Science*, 1980, *209*, 1140-1141.

Eimas, P., Siqueland, E. R., Jusczyk, P., & Vigorito, J. Speech perception in early infancy. *Science*, 1971, *171*, 304-306.

Elman, J. L. Perceptual origins of the phoneme boundary effect and selective adaptation to speech: A signal detection theory analysis. *Journal of the Acoustical Society of America*, 1979, *65*, 190-207.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. Perceptual equivalence of two acoustic cues for stop consonant manner. *Perception & Psychophysics*, 1980, *27*, 343-350.

Flege, J. E. The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly*, 1981, *15*, 443-455.

Flege, J. E., & Hammond, R. M. Mimicry of non-distinctive phonetic difference between language varieties. *Studies in Second Language Acquisition*, in press.

Fodor, J. A., Garrett, M. F., & Brill, S. L. Pi Ka Pu: The perception of speech sounds by prelinguistic infants. *Perception & Psychophysics*, 1975, *18*, 74-78.

Frazier, L. What can /w/, /l/, /y/ tell us about categorical perception? *Haskins Laboratories Status Report on Speech Research*, 1976, *SR-48*, 235-256.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. The identification and discrimination of synthetic vowels. *Language and Speech*, 1962, *5*, 171-189.

Fujisaki, H., & Kawashima, T. The influence of various factors on the identification and discrimination of synthetic speech sounds. *Reports of the 6th International Congress on Acoustics* (No. 2). Tokyo: 1968, pp. 95-98.

Fujisaki, H., & Kawashima, T. On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute* (Faculty of Engineering, University of Tokyo), 1969, *28*, 67-73.

Fujisaki, H., & Kawashima, T. Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute* (Faculty of Engineering, University of Tokyo), 1970, *29*, 207-214.

Fujisaki, H., & Kawashima, T. A model of the mechanisms for speech perception: Quantitative analysis of categorical effects in discrimination. *Annual Report of the Engineering Research Institute* (Faculty of Engineering, University of Tokyo), 1971, *30*, 59-68.

Fujisaki, H., & Shigeno, S. Context effects in the categorization of speech and non-speech stimuli. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers*. New York: Acoustical Society of America, 1979, pp. 5-8.

Ganong, W. F. III. *Selective adaptation and speech perception*. Unpublished doctoral dissertation, MIT, 1977.

Ganong, W. F. III. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, *6*, 110-125.

Garcia, E. The identification and discrimination of synthetic nasals. *Haskins Laboratories Status Report on Speech Research*, 1966, *SR-7/8*, 3.1-3.16

Garcia, E. Labelling of synthetic nasals (II). *Haskins Laboratories Status Report on Speech Research*, 1967, *SR-9*, 4.1-4.17. (a)

Garcia, E. Discrimination of three-formant nasal-vowel syllables (III). *Haskins Laboratories Status Report on Speech Research*, 1967, *SR-12*, 143-153. (b)

Gerstman, L. *Cues for distinguishing among fricatives, affricates, and stop consonants*. Unpublished doctoral dissertation, New York University, 1957.

Gibson, J. J. *The senses considered as perceptual systems*. Boston, Mass.: Houghton-Mifflin, 1966.

Godfrey, J. J. Perceptual difficulty and the right ear advantage for vowels. *Brain and Language*, 1974, *1*, 323-336.

Godfrey, J. J., & Millay, K. K. Discrimination of the "tempo of frequency change" cue. *Journal of the Acoustical Society of America*, 1981, *69*, 1446-1448.

Goto, H. Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia*, 1971, *9*, 317-323.

Grunke, M. E., & Pisoni, D. B. Some experiments on perceptual learning of mirror-image acoustic patterns. *Research on Speech Perception* (Progress Report No. 5). Bloomington, IN: Indiana University, Department of Psychology, 1979, pp. 147-182.

Haggard, M. P. Theoretical issues in speech perception. *Speech Synthesis and Perception* (Report No. 4). Cambridge, England: University of Cambridge, Psychology Laboratory, 1970, pp. 1-16.

Haggard, M., Summerfield, Q., & Roberts, M. Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading Fo cues in the voiced-voiceless distinction. *Journal of Phonetics*, 1981, *9*, 49-62.

Hanson, V. L. Within-category discrimination in speech perception. *Perception & Psychophysics*, 1977, *21*, 423-430.

Harris, J. D. Remarks on the determination of a differential threshold by the so-called ABX technique. *Journal of the Acoustical Society of America*, 1952, *24*, 417.

Harris, K. S. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1958, *1*, 1-7.

Harris, K. S., Bastian, J., & Liberman, A. M. Mimicry and the perception of a phonemic contrast induced by silent interval: Electromyographic and acoustic measures. *Journal of the Acoustical Society of America*, 1961, *33*, 842. (Abstract)

Hasegawa, A. *Some perceptual consequences of fricative coarticulation*. Unpublished doctoral dissertation, Purdue University, 1976.

Healy, A. F., & Levitt, A. G. Accessibility of the voicing distinction for learning phonological rules. *Memory & Cognition*, 1980, *8*, 107-114.

Healy, A. F., & Repp, B. H. Context independence and phonetic mediation in categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1982, *8*, 68-80.

Heider, E. R., & Olivier, D. C. The structure of the color space in naming and memory for two languages. *Cognitive Psychology*, 1972, *3*, 337-354.

Hillenbrand, J. Auditory perception experiments with sine-wave analogs to the voice-onset time dimension. *Journal of the Acoustical Society of America*, 1982, *71* (Suppl. No. 1), S74-75.

Hirsh, I. J. Auditory perception of temporal order. *Journal of the Acoustical Society of America*, 1959, *31*, 759-767.

Jusczyk, P. W. Perception of syllable-final stop consonants by 2-month-old infants. *Perception & Psychophysics*, 1977, *21*, 450-454.

Jusczyk, P. W. Infant speech perception: A critical appraisal. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives in the study of speech*. Hillsdale, N.J.: Erlbaum, 1981, pp. 113-164.

Jusczyk, P. On characterizing the development of speech perception. In J. Mehler & R. Fox (Eds.), *Neonate cognition: Beyond the blooming, buzzing confusion*. Hillsdale, N.J.: Erlbaum, in press.

Jusczyk, P. W., Copan, H., & Thompson, E. Perception by 2-month-old infants of glide contrasts in multisyllabic utterances. *Perception & Psychophysics*, 1978, *24*, 515-520.

Jusczyk, P. W., Pisoni, D. B., Walley, A., & Murray, J. Discrimination of relative onset time of two-component tones by infants. *Journal of the Acoustical Society of America*, 1980, *67*, 262-270.

Jusczyk, P. W., Rosner, B. S., Cutting, J. E., Foard, C. L., & Smith, L. B. Categorical perception of nonspeech sounds by 2-month-old infants. *Perception & Psychophysics*, 1977, *21*, 50-54.

Jusczyk, P. W., Smith, L. B., & Murphy, C. The perceptual classification of speech. *Perception & Psychophysics*, 1981, *30*, 10-23.

Jusczyk, P. W., & Thompson, E. Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception & Psychophysics*, 1978, *23*, 105-109.

Kaplan, H. L., Macmillan, N. A., & Creelman, C. D. Tables for d' from matching and oddity experiments. *Behavior Research Methods & Instrumentation*, 1978, *10*, 796-813.

Keating, P., & Blumstein, S. E. Effects of transition length on the perception of stop consonants. *Journal of the Acoustical Society of America*, 1978, *64*, 57-64.

Keating, P. A., Mikos, M. J., & Ganong, W. F. III. A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America*, 1981, *70*, 1261-1271.

Kewley-Port, D., & Pisoni, D. B. Discrimination of rise time in nonspeech signals: Is it categorical or noncategorical? *Journal of the Acoustical Society of America*, 1982, *71* (Suppl. No. 1), S36.

Kirstein, E. Perception of second-formant transitions in non-speech patterns. *Haskins Laboratories Status Report on Speech Research*, 1966, *SR-7/8*, 9.1-9.3.

Kopp, J., & Lane, H. Hue discrimination related to linguistic habits. *Studies in Language and Language Behavior* (Progress Report No. VI). Ann Arbor, MI: University of Michigan, 1968, 11-21.

Kopp, J., & Livermore, J. Differential discrimination or response bias? A signal detection analysis of categorical perception. *Journal of Experimental Psychology*, 1973, *101*, 179-182.

Kopp, J., & Udin, H. Identification and discrimination functions for pure tone frequencies. *Psychonomic Science*, 1969, *16*, 95-96.

Kuhl, P. K. Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*, 1979, *66*, 1668-1679. (a)

Kuhl, P. K. The perception of speech in early infancy. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice* (Vol. 1). New York: Academic Press, 1979, pp. 1-47. (b)

Kuhl, P. K. Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America*, 1981, *70*, 340-349.

Kuhl, P. K., & Miller, J. D. Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants. *Science*, 1975, *190*, 69-72.

Kuhl, P. K., & Miller, J. D. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, 1978, *63*, 905-917.

Kuhl, P. K., & Padden, D. M. *Speech perception by macaques: Enhanced discrimination at phonetic boundaries for place of articulation.* Manuscript submitted for publication, 1982. (a)

Kuhl, P. K., & Padden, D. M. *Speech discrimination by macaques: Potential auditory constraints on the evolution of language*. Manuscript submitted for publication, 1982. (b)

Kunisaki, O., & Fujisaki, H. On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* (University of Tokyo), 1977, *11*, 85-91.

Lane, H. L. Motor theory of speech perception: A critical review. *Psychological Review*, 1965, *72*, 275-309.

Lane, H. L. A behavioral basis for the polarity principle in linguistics. *Language*, 1967, *43*, 494-511.

Lane, H. L., & Moore, D. J. Reconditioning a consonant discrimination in an aphasic: An experimental case history. *Journal of Speech and Hearing Disorders*, 1962, *27*, 232-243.

Larkey, L. S., Wald, J., & Strange, W. Perception of synthetic nasal consonants in initial and final syllable position. *Perception & Psychophysics*, 1978, *23*, 299-312.

Lasky, R. E., Syrdal-Lasky, A., & Klein, R. E. VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*, 1975, *20*, 215-225.

Liberman, A. M. Some results of research on speech perception. *Journal of the Acoustical Society of America*, 1957, *29*, 117-123.

Liberman, A. M. On finding that speech is special. *American Psychologist*, 1982, *37*, 148-167.

Liberman, A. M., Cooper, F. S., Harris, K. S., MacNeilage, P. F., & Studdert-Kennedy, M. Some observations on a model for speech perception. In W. Wathen-Dunn (Ed.), *Models for the perception of speech and visual form*. Cambridge, Mass.: MIT Press, 1967, pp. 68-87.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. Psychological Review, 1967, 74, 431-461.

Liberman, A. M., Harris, K. S., Eimas, P. D., Lisker, L., & Bastian, J. An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. Language and Speech, 1961, 54, 175-195.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology, 1957, 53, 358-368.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. The discrimination of relative onset time of the components of certain speech and nonspeech patterns. Journal of Experimental Psychology, 1961, 61, 379-388.

Liberman, A. M., Mattingly, I. G., & Turvey, M. Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), Coding processes and human memory. Washington, D.C.: Winston, 1972, pp. 307-334.

Liberman, A. M., Studdert-Kennedy, M., Harris, K. S., & Cooper, F. S. A reply to "Identification and Discrimination Functions for a Visual Continuum and their Relation to the Motor Theory of Speech Perception" by Cross, Lane, and Sheppard. Haskins Laboratories Status Report on Speech Research, 1965, SR-3, 3.1.-3.14.

Lisker, L. On learning a new contrast. Haskins Laboratories Status Report on Speech Research, 1970, SR-24, 1-15.

Lisker, L. Stop voicing production: Natural outputs and synthesized inputs. Haskins Laboratories Status Report on Speech Research, 1976, SR-47, 155-164.

Lisker, L., & Abramson, A. A cross-language study of voicing in initial stops: Acoustical measurements. Word, 1964, 20, 384-422.

Lisker, L., & Abramson, A. S. The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences. Prague: Academia, 1970, pp. 563-567.

Lisker, L., Liberman, A. M., Erickson, D. M., Dechovitz, D., & Mandler, R. On pushing the voice-onset-time (VOT) boundary about. Language and Speech, 1977, 20, 209-216.

Locke, S., & Kellar, L. Categorical perception in a non-linguistic mode. Cortex, 1973, 9, 353-367.

MacKain, K. S., Best, C. T., & Strange, W. Categorical perception of English /r/ and /l/ by Japanese bilinguals. Applied Psycholinguistics, in press.

Macmillan, N. A., Kaplan, H. L., & Creelman, C. D. The psychophysics of categorical perception. Psychological Review, 1977, 84, 452-471.

Mandler, R. Categorical perception along an oral-nasal continuum. Haskins Laboratories Status Report on Speech Research, 1976, SR-47, 147-154.

Massaro, D. W. Language and information processing. In D. W. Massaro (Ed.), Understanding language. New York: Academic Press, 1975, pp. 3-28.

Mattingly, I. G., Liberman, A. M., Syrdal, A. M., & Halwes, T. Discrimination in speech and nonspeech modes. Cognitive Psychology, 1971, 2, 131-157.

May, J. G. Acoustic factors that may contribute to categorical perception. Language and Speech, 1981, 24, 273-284.

May, J. G., & Repp, B. H. Periodicity and auditory memory: A pilot study. Haskins Laboratories Status Report on Speech Research, 1982, SR-69, 145-149.

McClasky, C. L., Pisoni, D. B., & Carrell, T. D. Effects of transfer of training on identification of a new linguistic contrast in voicing. Research on Speech Perception (Progress Report No. 6). Bloomington, IN: Indiana University, Department of Psychology, 1980, pp. 205-234.

McGovern, K., & Strange, W. The perception of /r/ and /l/ in syllable-initial and syllable-final position. *Perception & Psychophysics*, 1977, *21*, 162-170.

McNabb, S. D. Must the output of phonetic feature detectors be binary? *Research on Speech Perception* (Progress Report No. 2). Bloomington, IN: Indiana University, Department of Psychology, 1976, pp. 166-179. (a)

McNabb, S. D. *Using confidence ratings to determine the sensitivity of phonetic feature detectors*. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1976. (b)

Mehler, J., & Bertoncini, J. Infants' perception of speech and other acoustic stimuli. In J. Morton & J. C. Marshall (Eds.), *Psycholinguistics 2: Structures and processes*. Cambridge: MIT Press, 1979, pp. 67-105.

Miller, J. D., Henderson, B. C., Sullivan, H. T., & Rigden, G. K. Speech perception by the chinchilla: Learning functions for pairs of synthetic stimuli from various points along a VOT continuum. *Journal of the Acoustical Society of America*, 1978, *64* (Suppl. No. 1), S18.

Miller, J. D., Wier, C. C., Pastore, R., Kelly, W. J., & Dooling, R. J. Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. *Journal of the Acoustical Society of America*, 1976, *60*, 410-417.

Miller, J. L. Properties of feature detectors for VOT: The voiceless channel of analysis. *Journal of the Acoustical Society of America*, 1977, *62*, 641-648.

Miller, J. L. Contextual effects in the discrimination of stop consonant and semivowel. *Perception & Psychophysics*, 1980, *28*, 93-95.

Miller, J. L. Phonetic perception: Evidence for context-dependent and context-independent processing. *Journal of the Acoustical Society of America*, 1981, *69*, 822-831.

Miller, J. L., & Connine, C. M. Psychophysical tuning curves for phonetically relevant acoustic information. *Journal of the Acoustical Society of America*, 1980, *67* (Supplement No. 1), S52. (Abstract)

Miller, J. L., & Eimas, P. D. Studies on the perception of place and manner of articulation: A comparison of the labial-alveolar and nasal-stop distinctions. *Journal of the Acoustical Society of America*, 1977, *61*, 835-845.

Miller, J. L., & Eimas, P. D. Biological constraints on the acquisition of language: Further evidence from the categorization of speech by infants. *Cognition*, in press.

Miller, J. L., Eimas, P. D., & Zatorre, R. J. Studies of place and manner of articulation in syllable-final position. *Journal of the Acoustical Society of America*, 1979, *66*, 1207-1210.

Miller, J. L., & Liberman, A. M. Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, 1979, *25*, 457-465.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, 1975, *18*, 331-340.

Molfese, D. L. Neuroelectrical correlates of categorical speech perception in adults. *Brain and Language*, 1978, *5*, 25-35.

Molfese, D. L. Hemispheric specialization for temporal information: Implications for the perception of voicing cues during speech perception. *Brain and Language*, 1980, *11*, 285-299.

Molfese, D. L., & Hess, T. M. Hemispheric specialization for VOT perception in the preschool child. *Journal of Experimental Child Psychology*, 1978, *26*, 71-84.

Molfese, D. L., & Molfese, V. J. VOT distinctions in infants: Learned or innate? In H. A. Whitaker & H. Whitaker (Eds.), *Studies in neurolinguistics (Vol.4)*. New York: Academic Press, 1979.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 1979, *7*, 323-331.

Morse, P. A. The infancy of infant speech perception: The first decade of research. *Brain, Behavior, and Evolution*, 1979, *16*, 351-373.

Morse, P. A., & Snowdon, C. T. An investigation of categorical speech discrimination by rhesus monkeys. *Perception & Psychophysics*, 1975, *17*, 9-16.

Neisser, U. *Cognition and reality*. San Francisco: Freeman, 1976.

Nusbaum, H. C., Schwab, E. C., & Sawusch, J. R. *Duplex perception: Dichotic integration or "chirp" identification*? Paper presented at the 52nd meeting of the Eastern Psychological Association, New York, April 1981.

Oden, G. C., & Massaro, D. W. Integration of featural information in speech perception. *Psychological Review*, 1978, *85*, 172-191.

Parks, T., Wall, C., & Bastian, J. Intercategory and intracategory discrimination for one visual continuum. *Journal of Experimental Psychology*, 1969, *81*, 241-245.

Pastore, R. E. Categorical perception: A critical re-evaluation. In S. K. Hirsh, D. H. Eldredge, I. J. Hirsh, & S. R. Silverman (Eds.), *Hearing and Davis: Essays honoring Hallowell Davis*. St. Louis, Mo.: Washington University Press, 1976, pp. 253-264.

Pastore, R. E. Possible psychoacoustic factors in speech perception. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives in the study of speech*. Hillsdale, N.J.: Erlbaum, 1981.

Pastore, R. E., Ahroon, W. A., Baffuto, K. J., Friedman, C., Puleo, J. S., & Fink, E. A. Common-factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 686-696.

Pastore, R. E., Friedman, C. J., & Baffuto, K. J. *A comparative evaluation of the AX and two ABX procedures*. Paper presented at the 92nd meeting of the Acoustical Society of America, San Diego, CA, November 1976.

Pastore, R. E., Harris, L. B., & Kaplan, J. K. Temporal order identification: Some parameter dependencies. *Journal of the Acoustical Society of America*, 1982, *71*, 430-436.

Perey, A. J., & Pisoni, D. B. *Dual processing vs. response-limitation accounts of categorical perception: A reply to Macmillan, Kaplan, and Creelman*. *Research on Speech Perception* (Progress Report No. 4). Bloomington, IN: Indiana University, Department of Psychology, 1978, pp. 197-216.

Perey, A. J., & Pisoni, D. B. Identification and discrimination of durations of silence in nonspeech signals. *Research on Speech Perception* (Progress Report No. 6). Bloomington, IN: Indiana University, Department of Psychology, 1980, pp. 235-269.

Pilch, H. Auditory phonetics. *Word*, 1979, *29*, 148-160.

Pisoni, D. B. *On the nature of categorical perception of speech sounds*. Unpublished doctoral dissertation, University of Michigan, 1971.

Pisoni, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 1973, *13*, 253-260.

Pisoni, D. B. The role of auditory short-term memory in vowel perception. *Memory & Cognition*, 1975, *3*, 7-18.

Pisoni, D. B. Discrimination of brief frequency glissandos. *Research on Speech Perception* (Progress Report No. 3). Bloomington, IN: Indiana University, Department of Psychology, 1976, pp. 155-169. (a)

Pisoni, D. B. Some effects of discrimination training on the identification and discrimination of rapid spectral changes. *Research on Speech Perception* (Progress Report No. 3). Bloomington, IN: Indiana University, Department of Psychology, 1976, pp. 122-141. (b)

Pisoni, D. B. Identification and discrimination of the relative onset of two component tones: Implications for the perception of voicing in stops. *Journal of the Acoustical Society of America*, 1977, *61*, 1352-1361.

Pisoni, D. B. Adaptation of the relative onset time of two-component tones. *Perception & Psychophysics*, 1980, *28*, 337-346. (a)

Pisoni, D. B. The perception of speech versus nonspeech. *Proceedings of the Ninth International Congress of Phonetic Sciences, Vol. III*. Copenhagen: University of Copenhagen, Institute of Phonetics, 1980, 301-312. (b)

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 1982, *8*, 297-314.

Pisoni, D. B., & Glanzman, D. L. Decision processes in speech discrimination as revealed by confidence ratings. *Research on Speech Perception* (Progress Report No. 1). Bloomington, IN: Indiana University, 1974, pp. 114-128.

Pisoni, D. B., & Lazarus, J. H. Categorical and non-categorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*, 1974, *55*, 328-333.

Pisoni, D. B., & Tash, J. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 1974, *15*, 285-290.

Pollack, I., & Pisoni, D. B. On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science*, 1971, *24*, 299-300.

Popper, R. D. Pair discrimination for a continuum of synthetic voiced stops with and without first and third formants. *Journal of Psycholinguistic Research*, 1972, *1*, 205-219.

Posner, M. I., & Mitchell, R. F. Chronometric analysis of classification. *Psychological Review*, 1967, *74*, 392-409.

Price, P. J. *A cross-linguistic study of flaps in Japanese and in American English*. Unpublished doctoral dissertation, University of Pennsylvania, 1981.

Raphael, L. J. Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English. *Journal of the Acoustical Society of America*, 1972, *51*, 1296-1303.

Raphael, L. J. Durations and contexts as cues to word-final cognate opposition in English. *Phonetica*, 1981, *38*, 126-147.

Raz, I., & Brandt, J. F. Categorical perception of nonspeech stimuli by musicians and nonmusicians. *Journal of the Acoustical Society of America*, 1977, *62* (Suppl. No. 1), S60. (Abstract)

Remez, R. E., Cutting, J. E., & Studdert-Kennedy, M. Cross-series adaptation using song and string. *Perception & Psychophysics*, 1980, *27*, 524-530.

Repp, B. H. *Categorical perception, dichotic interference, and auditory memory: A "same"-"different" reaction time study*. Unpublished manuscript, 1975. (Available from the author)

Repp, B. H. Identification of dichotic fusions. Journal of the Acoustical Society of America, 1976, 60, 456-469. (a)

Repp, B. H. "Posner's paradigm" and categorical perception: A negative study. Haskins Laboratories Status Report on Speech Research, 1976, SR-45/46, 153-161. (b)

Repp, B. H. Dichotic competition of speech sounds: The role of acoustic stimulus structure. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 37-50.

Repp, B. H. Perceptual equivalence of two kinds of ambiguous speech stimuli. Bulletin of the Psychonomic Society, 1981, 18, 12-14. (a)

Repp, B. H. Phonetic and auditory trading relations between acoustic cues in speech perception: Preliminary results. Haskins Laboratories Status Report on Speech Research, 1981, SR-67/68, 165-190. (b)

Repp, B. H. Two strategies in fricative discrimination. Perception & Psychophysics, 1981, 30, 217-227. (c)

Repp, B. H. Phonetic trading relations and context effects: New evidence for a phonetic mode of perception. Psychological Bulletin, in press.

Repp, B. H., Healy, A. F., & Crowder, R. G. Categories and context in the perception of isolated steady-state vowels. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 129-145.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. Perceptual integration of temporal cues for stop, fricative, and affricate manner. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 621-637.

Roberts, M., & Summerfield, Q. Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. Perception & Psychophysics, 1981, 30, 309-314.

Rosen, S. M. Range and frequency effects in consonant categorization. Journal of Phonetics, 1979, 7, 393-402.

Rosen, S. M., & Howell, P. Plucks and bows are not categorically perceived. Perception & Psychophysics, 1981, 30, 156-168.

Sachs, R. M. Vowel identification and discrimination in isolation vs. word context. Quarterly Progress Report No. 93. Cambridge, MA: MIT, Research Laboratory of Electronics, 1969, pp. 220-229.

Sachs, R. M., & Grant, K. W. Stimulus correlates in the perception of voice onset time (VOT): II. Discrimination of speech with high and low stimulus uncertainty. Journal of the Acoustical Society of America, 1976, 60 (Suppl. No. 1), S91. (Abstract)

Samuel, A. G. The effect of discrimination training on speech perception: Noncategorical perception. Perception & Psychophysics, 1977, 22, 321-330.

Samuel, A. G. Speech is specialized, not special. Unpublished doctoral dissertation, University of California at San Diego, 1979.

Samuel, A. G. Phonetic prototypes. Perception & Psychophysics, 1982, 31, 307-314.

Sawusch, J. R. Selective adaptation effects on end-point stimuli in a speech series. Perception & Psychophysics, 1976, 20, 61-65.

Sawusch, J. R., & Jusczyk, P. Adaptation and contrast in the perception of voicing. Journal of Experimental Psychology: Human Perception and Performance, 1981, 7, 408-421.

Sawusch, J. R., & Nusbaum, H. C. Contextual effects in vowel perception I: Anchor-induced contrast effects. Perception & Psychophysics, 1979, 25, 292-302.

Sawusch, J. R., Nusbaum, H. C., & Schwab, E. C. Contextual effects in vowel perception II: Evidence for two processing mechanisms. Perception & Psychophysics, 1980, 27, 421-434.

Schouten, M. E. H. The case against a speech mode of perception. Acta Psychologica, 1980, 44, 71-98.

Schwab, E. C. Auditory and phonetic processing for tone analogs of speech. Unpublished doctoral dissertation, SUNY at Buffalo, 1981.

Shigeno, S., & Fujisaki, H. Context effects in phonetic and non-phonetic vowel judgments. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics (University of Tokyo), 1980, 14, 217-224.

Siegel, J. A., & Siegel, W. Absolute identification of notes and intervals by musicians. Perception & Psychophysics, 1977, 21, 143-152. (a)

Siegel, J. A., & Siegel, W. Categorical perception of tonal intervals: Musicians can't tell sharp from flat. Perception & Psychophysics, 1977, 21, 399-407. (b)

Sinnott, J. M., Beecher, M. D., Moody, D. B., & Stebbins, W. C. Speech sound discrimination by monkeys and humans. Journal of the Acoustical Society of America, 1976, 60, 687-695.

Stevens, K. N. On the relations between speech movements and speech perception. Zeitschrift für Phonetik, Sprachwissenschaft, und Kommunikationsforschung, 1968, 21, 102-106.

Stevens, K. N., & Blumstein, S. E. Invariant cues for place of articulation in stop consonants. Journal of the Acoustical Society of America, 1978, 64, 1358-1368.

Stevens, K. N., & Klatt, D. H. Role of formant transitions in the voiced-voiceless distinction for stops. Journal of the Acoustical Society of America, 1974, 55, 653-659.

Stevens, K. N., Liberman, A. M., Öhman, S. E. G., & Studdert-Kennedy, M. Crosslanguage study of vowel perception. Language and Speech, 1969, 12, 1-23.

Strange, W. The effects of training on the perception of synthetic speech sounds: Voice onset time. Unpublished doctoral dissertation, University of Minnesota, 1972.

Strange, W., Edman, T. R., & Jenkins, J. J. Acoustic and phonological factors in vowel identification. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 643-656.

Strange, W., & Halwes, T. Confidence ratings in speech perception research: Evaluation of an efficient technique for discrimination testing. Perception & Psychophysics, 1971, 9, 182-186.

Strange, W., & Jenkins, J. J. Role of linguistic experience in the perception of speech. In R. D. Walk & H. L. Pick, Jr. (Eds.), Perception and experience. New York: Plenum Press, 1978, pp. 125-169.

Strange, W., Edman, T. R., & Jenkins, J. J. Acoustic and phonological factors in vowel identification. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 643-656.

Streeter, L. A. Language perception of 2-month old infants shows effects of both innate mechanisms and experience. Nature, 1976, 259, 39-41.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. Motor theory of speech perception: A reply to Lane's critical review. Psychological Review, 1970, 77, 234-249.

Studdert-Kennedy, M., Liberman, A. M., & Stevens, K. N. Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. *Journal of the Acoustical Society of America*, 1963, *35*, 1900. (Abstract)

Studdert-Kennedy, M., Liberman, A. M., & Stevens, K. N. Reaction time during the discrimination of synthetic stop consonants. *Journal of the Acoustical Society of America*, 1964, *36*, 1989. (Abstract)

Summerfield, Q. Use of visual information for phonetic perception. *Phonetica*, 1979, *36*, 314-331.

Summerfield, Q. Does VOT equal TOT or NOT? Examination of a possible auditory basis for the perception of voicing in initial stops. *Journal of the Acoustical Society of America*, in press.

Summerfield, Q., & Haggard, M. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, 1977, *62*, 436-448.

Swoboda, P., Kass, J., Morse, P., & Leavitt, L. Memory factors in infant vowel discrimination of normal and at risk infants. *Child Development*, 1978, *49*, 332-339.

Syrdal-Lasky, A. Effects of intensity on the categorical perception of stop consonants and isolated second formant transitions. *Perception & Psychophysics*, 1978, *23*, 420-432.

Tartter, V. C. A comparison of the identification and discrimination of synthetic vowel and stop consonant stimuli with various acoustic properties. *Journal of Phonetics*, 1981, *9*, 477-486.

Tartter, V. C. Vowel and consonant manipulations and the dual-coding model of auditory storage: A re-evaluation. *Journal of Phonetics*, 1982, *10*, 217-224.

Trehub, S. E. The discrimination of foreign speech contrasts by adults and infants. *Child Development*, 1976, *47*, 466-472.

Van Tasell, D. J., & Crump, E. S. A. Effects of stimulus level on perception of two acoustic cues in speech. *Journal of the Acoustical Society of America*, 1981, *70*, 1527-1529.

Van Valin, R. D., Jr. Perceived distance between vowel stimuli. *Journal of Phonetics*, 1976, *4*, 51-58.

Vinegrad, M. D. A direct magnitude scaling method to investigate categorical versus continuous modes of speech perception. *Language and Speech*, 1972, *15*, 114-121.

Walley, A. C., Pisoni, D. B., & Aslin, R. N. The role of early experience in the development of speech perception. In R. N. Aslin, J. Alberts, & M. R. Peterson (Eds.), *Sensory and perceptual development*. New York: Academic Press, 1981.

Wang, W. S.-Y. Language change. *Annals of the New York Academy of Science*, 1976, *280*, 61-72.

Waters, R. S., & Wilson, W. A., Jr. Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. *Perception & Psychophysics*, 1976, *19*, 285-289.

Watkins, M. J., & Todres, A. K. Suffix effects manifest and concealed: Further evidence for a 20-second echo. *Journal of Verbal Learning and Verbal Behavior*, 1980, *19*, 46-53.

Werker, J. F. Personal communication, March 1982.

Werker, J. F. *The development of cross-language speech perception: The effect of age, experience, and context on perceptual organization*. Unpublished doctoral dissertation, University of British Columbia, Vancouver, B.C., 1982.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. Developmental aspects of cross-language speech perception. Child Development, 1981, 52, 349-355.

Williams, L. The perception of stop consonant voicing by Spanish-English bilinguals. Perception & Psychophysics, 1977, 21, 289-297.

Wood, C. C. Discriminability, response bias, and phoneme categories in discrimination of voice onset time. Journal of the Acoustical Society of America, 1976, 60, 1381-1389. (a)

Wood, C. C. Discriminability, response bias, and phoneme categories in speech discrimination: II. Place of articulation. Paper presented at the 91st Meeting of the Acoustical Society of America, Washington, D.C., April 1976. (b)

Zatorre, R. J., & Halpern, A. R. Identification, discrimination, and selective adaptation of simultaneous musical intervals. Perception & Psychophysics, 1979, 26, 384-395.

SHORT-TERM RECALL BY DEAF SIGNERS OF AMERICAN SIGN LANGUAGE: IMPLICATIONS OF ENCODING STRATEGY FOR ORDER RECALL*

Vicki L. Hanson

Abstract. Two experiments were conducted on short-term recall of printed English words by deaf signers of American Sign Language (ASL). Compared with hearing subjects, deaf subjects recalled significantly fewer words when ordered recall of words was required, but not when free recall was required. Deaf subjects tended to use a speech-based code in probed recall for order and the greater the reliance on a speech-based code, the more accurate the recall. These results are consistent with the hypothesis that a speech code facilitates the retention of order information.

For hearing persons, short-term retention of English letters and words tends to employ a speech-based code. This is true regardless of whether the input items are spoken (Baddeley, 1966; Hintzman, 1967; Wickelgren, 1965, 1966) or written (Conrad, 1962, 1964; Kintsch & Buschke, 1969; Posner, Boies, Eichelman, & Taylor, 1969). It has been hypothesized that not only may this speech-based code not only be well suited for representing linguistic material in short-term memory, but that it may also be particularly well suited for retention of order information (Baddeley, 1978; Crowder, 1978; Healy, 1975). Whether or not there are properties of a speech-based code that make it particularly effective for short-term retention of words can be tested by examining short-term recall by congenitally and profoundly deaf signers of American Sign Language (ASL).

ASL, the visual-gestural language used in deaf communities in North America, is acquired by children of deaf parents as a native language. It

---

*To appear in Journal of Experimental Psychology: Learning, Memory and Cognition.

Acknowledgment. I am grateful to Ben Bahan, Nancy Fishbein, Nancy Frishberg, and Dennis Schemenauer for their help in making arrangements for subjects to participate in the experiments, and to Venita Lutes-Driscoll, Maxine Schuster, Malinda Williams, and Ira Rothenberg for testing subjects, data transcription, and signing experimental tapes. I am also grateful for the cooperation of the following institutions in this research endeavor: Gallaudet College, New York University, and California State University, Northridge. Illustrations were drawn by Frank A. Paul. Valuable comments on earlier versions of this manuscript were provided by John Richards, Mike Shand, Laurie Feldman, Nancy McGarr, and Ursula Bellugi. This research was supported by the National Institute of Education Grant #NIE-G-80-0178 and by NINCDS Research Service Award #NS-06109, NICHD Grant HD-01994, and NINCDS grant NS-18010. Portions of this work were conducted while the author was a postdoctoral fellow at The Salk Institute.

differs from English not only in the grammatical structure of sentences (Klima & Bellugi, 1979), but also in the form of lexical structure. In spoken languages, word structure is based on sequential production of phonemes. In ASL, sign structure is based on the simultaneous production of the formational parameters of handshape, movement, and place of articulation (Stokoe, Casterline, & Croneberg, 1965). These formational parameters have no direct correspondence to English phonemes or letters (graphemes).

For deaf signers of ASL, short-term retention of signs has been found to use, not a speech-based code, but rather a sign-based code. Bellugi, Klima, and Siple (1975) have shown that intrusion errors in recall of signs are related to the formational parameters of the signs. For example, an intrusion error for deaf subjects on recall of the sign VOTE was tea, a word whose corresponding sign is similar in handshape and place of articulation to the sign VOTE, but differs in movement. Additional evidence for sign-based encoding of signs has been obtained by Frumkin and Anisfeld (1977) and by Poizner, Bellugi, and Tweney (1981).

Other work has been concerned with whether sign-based encoding is used by deaf persons in the short-term retention of printed English words. Odom, Blanton, and McIntyre (1970) presented deaf children (mean age 16.0 years) with lists of written words to learn. They compared the learning of a list of words having close sign correspondences with the learning of a list of "unsignable" words and found that the deaf children learned the list of signable words more easily than the list of "unsignable" words. The implication from these results is that the deaf children were recoding into a sign-based code when possible. Similar to their findings, Conlin and Paivio (1975), in a paired associate task, found that deaf high school and college students learned signable pairs of words more readily than pairs of words for which there were no direct sign translations. Moulton and Beasley (1975) found that their deaf subjects (mean age 18.0) learned pairs of words having formationally similar signs more readily than they learned pairs of words having formationally dissimilar signs. Shand (1982), testing adult signers in an ordered recall task, provided a test of speech-based as well as sign-based encoding of words. He found that lists of words having formationally similar signs were not as well recalled as were lists of words having formationally dissimilar signs. This finding was consistent with earlier work indicating the use of sign-based encoding. Lists of phonetically similar words, however, were not recalled less accurately by deaf signers than were lists of unrelated words, suggesting that speech-based encoding was not being used by the subjects.

The studies just summarized indicate that a sign-based code can be used as a basis for representing linguistic material in short-term memory, but are unanalytic with respect to the question of whether there are special properties of a speech-based or sign-based code that might make a particular encoding strategy most effective on a given task. The present experiments provide such an examination as it relates to one hypothesized function of a speech-based code: retention of order information (Baddeley, 1978; Crowder, 1978; Healy, 1975). This study investigates speech-based and sign-based encoding of printed words by deaf native signers of ASL. Two experiments are reported here. The first is an ordered recall paradigm, requiring recall of items and the order in which they are presented; the second is a free recall paradigm, requiring recall of items regardless of order. If temporal order information is most effectively retained by a speech-based code, then persons not using this code should be hindered in the ordered recall task of Experiment 1. If retention of item information, however, does not require the use of a speech-based code, then recall accuracy should not be related to the use of a speech-based code in the free recall task of Experiment 2.

EXPERIMENT 1

In Experiment 1, the encoding of printed words by deaf native signers of ASL was investigated using a modified version of the ordered recall paradigm developed by Baddeley (1966). The paradigm involves presentation of sets of words chosen to be similar along one dimension. Each similar set is matched with a control set of words that bear no similarity to each other. With spoken word presentations, Baddeley found that for hearing persons there is a decrement in performance when spoken words to-be-recalled are phonetically similar. Using this paradigm with ASL sign presentations, Poizner et al. (1981) found that for deaf signers there is a decrement in performance when signs-to-be-recalled are formationally similar.

METHOD

Stimulus Sets

Three experimental sets of eight monosyllabic words each were constructed: 1) formationally (sign) similar, 2) phonetically similar, and 3) graphemically similar. For each of these three experimental sets, a control set of words was constructed. Each control set was matched with its corresponding experimental set for part of speech and for frequency of occurrence in written English (Thorndike & Lorge, 1944). As a result, performance on an experimental set is only interpretable in relation to performance on its matched control. A practice set, consisting of words unrelated to each other, was also constructed. Deaf signers (not participating in the experiment) acted as ASL informants regarding the corresponding signs for each English word.

The words in the formationally similar set were phonetically and graphemically dissimilar. The criteria for formational similarity were that the signs for each of the words were signed with similar handshapes and with the two hands contacting in neutral space in front of the body. The following eight words were, as a result, selected: KNIFE, EGG, NAME, PLUG, TRAIN, CHAIR, TENT, SALT. Illustrations of signs for these words appear in Figure 1.

The words of the phonetically similar set rhymed and were formationally and graphemically dissimilar as possible. The eight words of the phonetically similar set were the following: TWO, BLUE, WHO, CHEW, SHOE, THROUGH, JEW, YOU. Since some graphemic similarity was unavoidable for this set, an experimental set of graphemically similar words was constructed to tease apart possible confounding effects due to this similarity. The following words were used for this latter set: BEAR, MEAT, HEAD, YEAR, LEARN, PEACE, BREAK, DREAM.

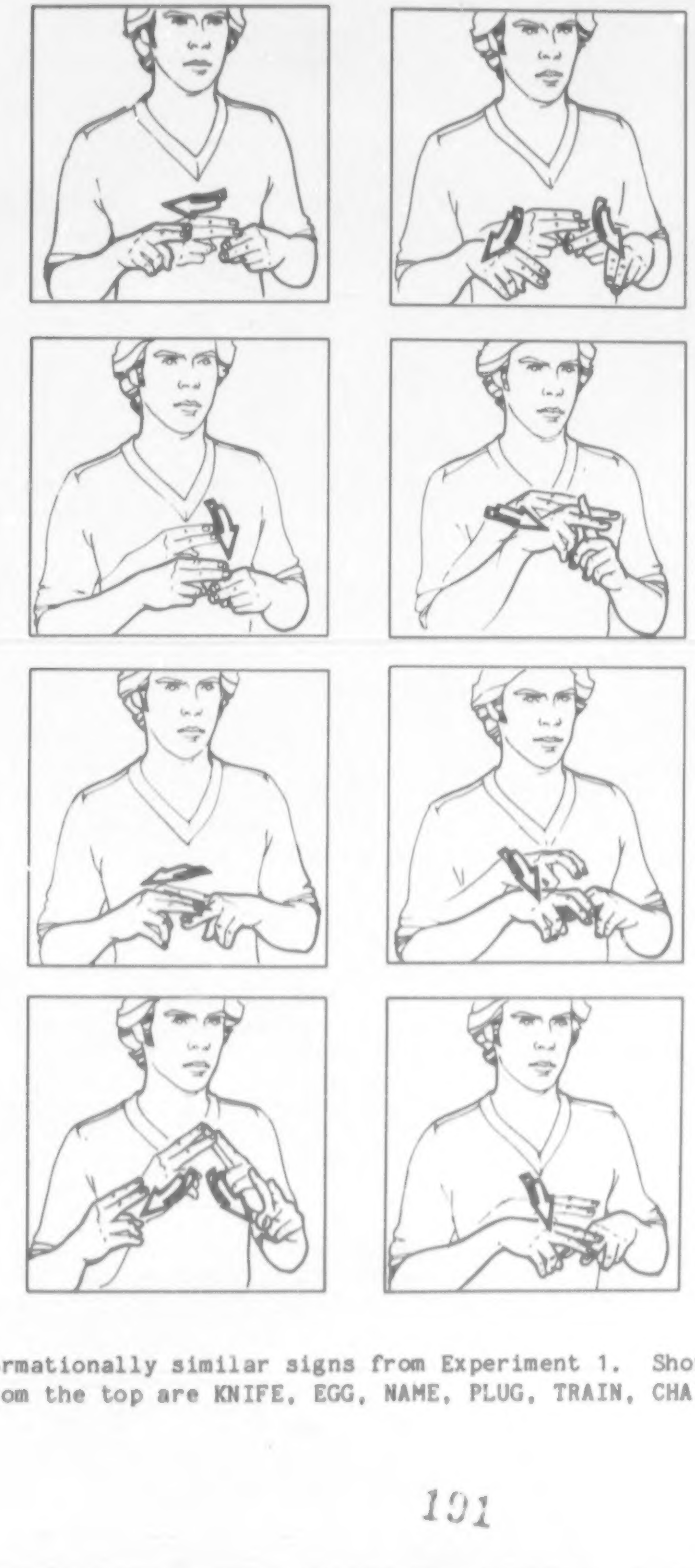


Figure 1. Formationally similar signs from Experiment 1. Shown left to right from the top are KNIFE, EGG, NAME, PLUG, TRAIN, CHAIR, TENT, SALT.

Appendix A lists all the words for the experimental and control sets.

Design

A group of hearing subjects and a group of deaf subjects were tested with the printed words. To ensure that the stimuli were appropriate for detecting sign encoding, an additional condition was run. As previous work has shown that sign presentation elicits sign-based encoding of the stimuli (Bellugi et al., 1975; Poizner et al., 1981), a second group of deaf subjects was tested with signed presentation of the stimulus items.

Procedure

The paradigm of Baddeley (1966) was modified here to be a probed recall task. In this task, a series of five words (or signs) was presented, followed by a probe (one of the first four of the just-presented items). Subjects responded by indicating the word (or sign) following that probe in the series.

Printed word presentation. A micro-computer was used for stimulus display and data collection. Trials were blocked by stimulus set. The order of experimental set presentation was randomized, with the restriction that an experimental set and its control were always presented consecutively. Prior to testing with each set, the eight words of the set were displayed. The words were each assigned a number (1-8) and the word and its number were typed on 3" x 5" index cards. This card was continuously displayed during the 16 trials of testing with a set.

On each trial, subjects were presented a warning signal, a "+", followed by five words consecutively displayed in the center of the CRT screen. The words were printed in all upper-case letters and were shown at a rate of one second per word. Word order was random with the constraint that each word appeared twice in each serial position during a block. Each of the eight words of a set was used twice as a probe word and twice as an answer.

The probe word was presented three sec after the last stimulus word. Subjects responded with the word that followed the probe on that trial, pressing the key on the computer terminal indicating the number of the word that was their answer. This response procedure was chosen for two reasons. First it was necessary to provide a response that could be used equally well by deaf and hearing subjects. Second, pilot testing had indicated that writing the words tended to encourage many deaf subjects to fingerspell as they were writing. Fingerspelling is a system based on English in which there is a manual configuration for each letter of the alphabet and words are spelled by the sequential production of each letter. Due to the similarity of spellings for the words in the graphemically similar list, it was desirable not to use a response procedure that would specifically encourage such a strategy.

Instructions were written. Additionally, a summary of the instructions was signed for deaf subjects and spoken for hearing subjects.

Sign presentation. The signed stimuli were recorded on videotape by a native signer of ASL at the same rate of presentation used with the printed words. The signer maintained a neutral expression throughout the signing of the stimuli (i.e., no mouth movement nor facial expressions accompanied the signs). Instructions, signed in ASL, were recorded on the beginning of the test videotape.

Contraints imposed by the use of videotaped rather than computer-displayed stimuli necessitated a few procedural differences from the printed word condition. Rather than having the card with the English words presented during a block, subjects were given a paper on which the signs for that block were drawn as in Figure 1. Subjects responded by signing the item that followed the probe. A videotape was made of each subject in this sign presentation condition, and the videotaped answers of each subject were later transcribed. Stimulus sets were presented to subjects in the following fixed order: practice set, formational control set, formationally similar set, phonetically similar set, phonetic control set, graphemically similar set, and graphemic control set.

Subjects

Three groups of subjects were tested. They were paid for their participation in the experiment, which lasted approximately one hour.

Sign presentation. Seven prelingually deaf volunteers were recruited through the Salk Institute and through California State University, Northridge. Five had a hearing loss of 90 dB or greater in the better ear. The remaining two subjects had a loss of 70 dB in the better ear. All were native signers of ASL.

Printed word presentation. Hearing subjects were eight college-age persons who responded to an ad in a local paper requesting subjects for a psychology experiment.

Deaf subjects were eight volunteers recruited through The Salk Institute, California State University, Northridge, and Gallaudet College. All were native signers of ASL. Two were recent college graduates and the other six were presently enrolled in college. With only one exception, deaf subjects had a hearing loss of 90 dB or greater in the better ear. That one subject had a loss of 80 dB in the better ear.

RESULTS AND DISCUSSION

Encoding

Sign presentation. Data from the sign presentation condition were examined to determine whether the stimulus materials were suitable for detecting sign encoding. A deaf native signer of ASL assisted in the transcription of the signed responses. Subjects were found to be significantly less accurate on the formationally similar set than on the formational control set, $t(6)=4.19$, $p<.01$. This significant decrement for the formationally similar set is in agreement with other work indicating sign-based encoding when ASL signs are presented (Bellugi et al., 1975; Poizner et al., 1981). For purposes of the present study, it demonstrates that the formationally similar set was appropriate for detecting sign encoding. Results are given in Table 1.

Table 1

Percentage Correct Trials for Each Stimulus Set in Experiment 1.

	Formational	Phonetic	Graphemic	Mean
Sign (Deaf)				
Similar	41.3	60.0	63.6	
Control	59.0	71.6	69.9	66.8
(Percentage Decrement)	17.7*	11.6*	6.3	
Deaf				
Similar	51.4	47.6	47.6	
Control	52.9	65.4	52.2	56.8
(Percentage Decrement)	1.5	17.8*	4.6	
Hearing				
Similar	87.4	70.2	86.7	
Control	84.2	96.9	89.9	90.3
(Percentage Decrement)	-3.2	26.7*	3.2	

(* $p<.05$)

Compared with its matched control, the graphemically similar set did not produce a significant decrement in performance, $t(6)=.75$, $p>.20$. An effect of phonetic similarity was found, however, with subjects being less accurate on the phonetically similar set than on its matched control set, $t(6)=3.15$, $p<.05$. This result is consistent with observation of subjects' rehearsal strategies on the recorded videotapes: rehearsal often involved the simultaneous signing and mouthing of the English word for each of the presented signs. This speech-based rehearsal occurred despite the neutral facial expression maintained by the signer during presentation of the signed stimuli.

Printed word presentation. For the printed words, an analysis of variance was performed on subject group (deaf vs. hearing) by dimension (formational, phonetic, vs. graphemic) by set (experimental set vs. control set). The analysis revealed an interaction of dimension by set, $F(2,28)=8.04$, $MS_e=146.96$, $p<.005$, indicating a significant decrement in performance only for some of the experimental sets. This effect did not significantly interact with group, $F(2,28)=.68$, $MS_e=146.96$, $p>.20$, suggesting a similar pattern of results for both deaf and hearing subjects. The percentage correct for the two groups on each set are given in Table 1.

Post hoc analyses on the simple effects revealed that subjects did not exhibit a significant performance decrement for the formationally similar set, $F(1,28)=.04$, $p>.20$. The subjects did, however, show a performance decrement for the phonetically similar set compared with its control set, $F(1,28)=26.80$, $p<.001$. There was no significant effect of graphemic similarity, $F(1,28)=.82$, $p>.20$, indicating that the decrement for the phonetically similar set was not due to graphemic similarity.

Since the sign presentation condition obtained evidence for sign-based encoding, it does not seem that the failure to find such evidence with printed words can be attributed to inappropriate stimulus materials or design. As the sign correspondence for each word in the formationally similar set is quite straightforward, it does not seem that failure to find evidence of sign-based encoding is attributable to variability in the word to sign translations. Rather, it appears that stimulus input had an effect on encoding strategy of deaf subjects: Presentation of ASL signs encouraged the use of sign-based encoding.

The present experiment suggests the use of speech-based encoding in short-term ordered recall by deaf adults. Both with sign and printed word presentation, subjects evidenced speech-based encoding. The reason for this cannot definitely be determined here, but it may be that speech-based rehearsal was in use due to the experimental situation. Given the requirement of order recall in the present experiment, subjects may have been influenced to use speech-based encoding.

Accuracy

The measure of overall accuracy in this experiment was the accuracy on the three control sets. With printed word presentation, the hearing subjects responded correctly significantly more often than did deaf subjects, $t(14)=4.53$, $p<.001$. This finding that deaf subjects had difficulty with ordered recall is consistent with other studies (Conrad, 1970; MacDougall, 1979; Pintner & Paterson, 1917; Wallace & Corballis, 1973) that have found poorer performance of deaf than hearing subjects on short-term memory tasks.

The difficulties of deaf populations on memory tasks has been often attributed to difficulties with English (Belmont & Karchmer, 1978; Furth, 1971). But work by Conrad (1979) suggests another interpretation. He found that memory span was related to use of phonetic coding. Those deaf subjects who used a speech-based code recalled more items in an ordered recall task than did those deaf subjects not using this code. It appeared, as a result, that recall accuracy in ordered recall was a function of speech encoding. Indeed, there is a similar suggestion from the present experiment. For the eight deaf subjects tested on recall of printed words, number of correct responses on the three control sets correlated with the performance decrement on the phonetically similar set, $r=.63$. That is, the larger the decrement due to phonetic similarity, and thus the greater the evidence for use of a speech-based code, the greater the recall accuracy for the subject. This suggests that recall accuracy in this ordered recall task may be a function of the use of a speech-based code.

EXPERIMENT 2

Experiment 2 was designed to address whether or not difficulties of deaf subjects in short-term recall are limited to ordered recall. The hypothesis that a speech-based code is particularly suitable for temporal order recall (Baddeley, 1978; Crowder, 1978; Healy, 1975) leads to the prediction that ordered recall should be difficult for persons not having normal access to speech input. Experiment 2 employed a free recall paradigm. If order recall, more than item recall, is dependent on the use of a speech-based code, then deaf subjects may not show short-term memory difficulties when only item recall is required.

Two conditions were included in Experiment 2: formational similarity and phonetic similarity. With hearing adults, Watkins, Watkins, and Crowder (1974) found that for free recall phonetic similarity of words in a list improved recall accuracy when compared with lists of unrelated words. Thus, when memory for order was not required, the phonetic similarity of words proved to be of benefit to subjects using a speech-based code. The phonetic similarity condition of the present experiment was similar to that of Watkins et al. (1974). Lists of phonetically similar words were constructed such that, compared with performance on unrelated lists of words, subjects using speech-based encoding should benefit from the phonetic similarity. In the formationally similar condition, lists of words were constructed such that the corresponding signs were formationally similar. Compared with performance on unrelated lists of words, formational similarity should improve performance if subjects are using sign-based encoding.

METHOD

Stimulus Sets

The formational similarity condition and the phonetic similarity condition each employed five sets of words. Each set contained an experimental list of formationally or phonetically similar words and a control list of unrelated words. There were 12 words per list. As in Experiment 1, words were chosen so that each English word had a corresponding sign.

For the formational similarity condition, each word in an experimental list had a corresponding sign that was formationally similar to the signs of the other words in the list. The signs for all words in the experimental lists were produced with both hands having the same handshape and with the place of articulation being neutral space in front of the body. For each of the five formationally similar lists, a different handshape was used. Each formationally similar list was matched with a control list for number of syllables and frequency of occurrence in written English (Thorndike & Lorge, 1944); thus, as in Experiment 1, performance on an experimental list was only interpretable in relation to performance on the matched control. The signs for words in each of the control lists were formationally dissimilar.

For the phonetic similarity condition, five lists of phonetically similar words were constructed. Each phonetically similar list was composed of monosyllabic words sharing the vowel sound. As much as possible, words in the phonetically similar lists were graphemically dissimilar. Control lists, matched as described above, were constructed for each of the phonetically similar lists.

Appendix B lists the sets of words.

Design

Four groups of subjects participated, a group of deaf subjects and a group of hearing subjects in each of the two conditions. To test whether the lists of words having formationally similar signs were suitable for obtaining evidence of sign encoding, an additional group of deaf subjects was tested. This group was instructed to think of the signs for each word presented in the formationally similar condition.

Procedure

A videotaped CRT display presented the twelve words of a list at the rate of one word every two seconds. All words were displayed in the center of the screen. The list presentation was followed by the instruction "WRITE ALL THE WORDS YOU REMEMBER." Subjects were given as much time as necessary to write their answers. Presentation of the next list then began. Each list presentation was preceded by the word "READY" displayed for two seconds.

A practice list was first presented followed by a random presentation of the ten test lists. Two different random list orders were used and half of the subjects were tested with each list order.

Instructions, signed in ASL, were also recorded on videotape. The instructions informed subjects that they would see several groups of twelve words. They were told that when they were given the recall cue they were to write all the words they could remember in any order they wanted. In the instructed condition, subjects were additionally told to think of the signs for the words presented and use the signs to help them recall the words. They were not, however, informed about the nature of the list construction.

Subjects

Subjects were tested in groups of one to three persons. They were paid for their participation in this 1/2 hour experiment.

Hearing subjects. Each group of hearing subjects was composed of eight staff members of The Salk Institute.

Deaf subjects. Deaf subjects were native signers of ASL recruited through The Salk Institute and California State University, Northridge, and through Gallaudet College. All had a hearing loss of 90 dB or greater in the better ear. All were currently enrolled in college or were recent college graduates. There were eight deaf subjects in each of the three groups.

RESULTS AND DISCUSSION

Encoding

To examine whether the formationally similar sets were suitable for obtaining evidence of sign encoding, the responses of the group instructed to use signs were analyzed in an analysis of variance for list type (experimental vs. control) by stimulus set (Sets 1 - 5). The results indicated no significant overall benefit due to formational similarity, $F(1,7)=1.90$, $MS_e=334.25$, $p>.20$, but there was a significant interaction of list type by set, $F(4,28)=4.52$, $MS_e=140.67$, $p<.01$. This indicated that benefit due to formational similarity was obtained only for some of the stimulus sets. Analysis of the simple effects revealed that only two of the five formationally similar lists showed a reliable improvement in performance compared with their matched control: Set 1, $F(1,28)=16.34$, $p<.001$; Set 2, $F(1,28)=5.19$, $p<.05$. For the other three sets, subjects actually recalled somewhat fewer words on the experimental list than on the control, although the differences were not significant: Set 3, $F(1,28)=.27$; Set 4, $F(1,28)=.03$; Set 5, $F(1,28)=.75$; all $p>.20$. While it is puzzling that the benefit due to formational similarity was not more generally obtained, suggesting that the sign analog of phonetic similarity was not completely captured in the present design of experimental stimuli, there were at least two sets of stimuli that were suitable for testing whether sign-based encoding is used in the task. Results shown in Table 2 indicate the benefit in performance due to formational similarity both for these two sets and for all sets.

Table 2

Percentage Correct Trials in Experiment 2.

	Sets 1 and 2	All Sets		
	Formational	Formational	Phonetic	Mean
Instructed (Deaf)				
Similar	66.1	60.0		
Control	47.4	54.4		
(Percentage Benefit)	18.7*	5.6		
Deaf				
Similar	51.0	54.2	50.2	
Control	46.4	49.8	47.1	48.4
(Percentage Benefit)	4.6	4.4	3.1	
Hearing				
Similar	55.7	55.8	53.5	
Control	56.3	56.5	42.5	49.5
(Percentage Benefit)	-.4	-.7	11.0*	

(* $p<.05$)

Analyses for the formational similarity condition were based only on those two sets of the formational similarity condition that appeared appropriate for obtaining evidence of sign-based encoding. An ANOVA was performed on percent correct for subject group (instructed [deaf], deaf, vs. hearing) by list type by stimulus set (Sets 1 and 2). The analysis revealed an overall benefit due to formational similarity, $\underline{F}(1,21)=4.82$, $\underline{MS}_e=290.94$, $\underline{p}<.05$, that tended to interact with subject group, $\underline{F}(2,21)=2.72$, $\underline{MS}_e=290.94$, $\underline{p}<.10$. Analysis of the simple effects revealed that there was a significant benefit due to formational similarity for deaf subjects in the instructed condition, $\underline{F}(1,21)=9.66$, $\underline{p}<.01$, but that the deaf subjects in the experimental group did not show a significant benefit due to formational similarity, $\underline{F}(1,21)=.60$, $\underline{p}>.20$. Hearing subjects, as expected, showed no benefit due to formational similarity, $\underline{F}(1,21)=.01$, $\underline{p}>.20$. This suggests that the deaf subjects, unless specifically instructed to do so, were not encoding the written words in terms of a sign-based code and is in accord with the results of Experiment 1 where sign-based encoding of printed words was not indicated.

So few intrusion errors were made on Sets 1 and 2 that analysis of the types of intrusions made was not feasible. In the instructed condition, deaf subjects made a total of 13 intrusions, 5 of which were in the formationally similar lists. Deaf subjects in the experimental group made 17 intrusion errors, 7 of which were made on recall of the formationally similar lists. Hearing subjects made 13 intrusion errors, 6 of which occurred on recall of the formationally similar lists.

The percentage correct for deaf and hearing subjects in the phonetic similarity condition was analyzed for group (deaf vs. hearing) by list type by set. Results indicated that there was a main effect of similarity, $\underline{F}(1,14)=21.09$, $\underline{MS}_e=95.08$, $\underline{p}<.001$, suggesting a benefit due to phonetic similarity. This effect interacted with group, however, $\underline{F}(1,14)=6.59$, $\underline{MS}_e=95.08$, $\underline{p}<.05$. Analysis of the simple effects indicated a significant benefit due to phonetic similarity for the hearing subjects, $\underline{F}(1,14)=25.63$, $\underline{p}<.001$, but not for the deaf subjects, $\underline{F}(1,14)=2.05$, $\underline{p}>.10$. The benefit of phonetic similarity for the hearing subjects did not interact with set, $\underline{F}(4,28)=.94$, $\underline{MS}_e=99.23$, $\underline{p}>.20$, reflecting benefit for all five stimulus sets.

Consistent with this finding, examination of the intrusion errors on the five sets revealed that hearing subjects, more often than deaf subjects, made intrusion errors consistent with the phonetically similar lists. Hearing subjects made a total of 33 intrusions. Of the 16 on the phonetically similar lists, 12 errors (75%) were phonetically similar to the other words. Deaf subjects made 36 intrusions, and of the 15 intrusions on the phonetically similar lists, only 2 errors (13%) were phonetically similar to the other words.

This experiment, then, was suitable for obtaining evidence of speech-based encoding, as the results of the hearing subjects indicate. However, evidence for the use of speech-based encoding by deaf subjects was not indicated. This would seem inconsistent with the results of Experiment 1 in which speech-based encoding was indicated. But rather than considering these results as inconsistent, two qualifying factors must be taken into account. The first is the task requirements. The task varied in the two experiments and this may have influenced encoding strategies.

The second factor to consider is that failure to find evidence of speech-based encoding by deaf subjects must be viewed with caution in studies relying on phonetic similarity for such detection. In these studies, no evidence of speech-based encoding will be obtained if subjects are using pronunciations different from those anticipated by the experimenter. As deaf adults at times differ from hearing adults in their judgments about whether or not pairs of printed words rhyme (Hanson, 1980), word lists constructed by the experimenter to be phonetically similar may not always be phonetically similar as pronounced by deaf subjects.

This caution applies to the interpretation of the present nonsignificant results for deaf subjects in the phonetic similarity condition. In this regard, it is worth examining the performance of deaf subjects on Set 1 in the phonetic similarity condition of Experiment 2. The experimental list of Set 1 contained words from the phonetically similar set of Experiment 1. In Experiment 1, these words did provide evidence of speech-based encoding, implying that subjects were using the expected pronunciations of words. It is interesting to note that for Set 1, deaf subjects in Experiment 2 did recall more words from the experimental list than from its control, $\underline{t}(7)=2.88$, $\underline{p}<.05$. While it would be inappropriate to draw strong conclusions from this analysis, it is interesting to note that the finding is consistent with the hypothesis that failure to find evidence of speech-based encoding may result, at least in part, from deaf subjects not using the expected pronunciations of words.

Accuracy

Of concern in the present study is overall accuracy in the free recall task of Experiment 2. To address this issue, the percentage correct for all control lists was analyzed. The ANOVA on data from the four experimental groups indicated that there was no significant difference in recall accuracy for deaf and hearing subjects, $\underline{F}(1,28)=.07$, $\underline{MS}_e=583.12$, $\underline{p}>.20$. This finding is of major interest since memory studies typically show performance levels of deaf subjects to be lower than performance levels of hearing subjects (Conrad, 1970; MacDougall, 1979; Wallace & Corballis, 1973). The comparable recall accuracy of deaf and hearing subjects in this free recall task was also in marked contrast to the results of the ordered recall task used in Experiment 1.

In a search of the literature, only one previous study was found that was concerned with free recall accuracy of words by deaf subjects. In that research, by Koh, Vernon, and Bailey (1971), it was found that deaf subjects recalled about one item less than hearing subjects did. However, a methodological confounding noted by the authors makes it uncertain whether their study actually tested memory for words. In the method employed, pictures of each of the words were presented simultaneously with the written words, perhaps influencing subjects toward use of memory strategies different from those employed in recall of purely linguistic material.

In the present task, then, which required only item recall, deaf subjects were not found to have short-term memory deficits as compared with hearing subjects. This finding raises the question of how the item information was retained, as evidence was obtained for use of neither a speech-based nor a sign-based code by deaf subjects. With hearing subjects, Healy (1977) found evidence indicating a non-speech code involved in retention of item information. It is not unreasonable to expect that deaf subjects might make extensive use of this (perhaps visual) code in recall of item information. However, the above caution regarding failure to find evidence of speech-based coding by deaf subjects must be borne in mind before concluding that deaf subjects were not employing such a code in Experiment 2.

GENERAL DISCUSSION

In understanding the nature of the internal representation of English words for deaf persons, it may be necessary to discuss encoding as it relates to specific subjects in a specific task rather than trying to determine the encoding strategy employed by deaf persons. The present research is consistent with earlier research in finding that adult signers are able to use a sign-based code for short-term retention of linguistic material (Bellugi et al., 1975; Conlin & Paivio, 1975; Poizner et al., 1981; Shand, 1982), although the present findings further suggest that factors such as stimulus input (signs or printed English words) and task requirements are likely to influence encoding strategy. Although not examined in the present research, individual subject characteristics such as degree of hearing loss, linguistic background, access to a speech-based code (Conrad, 1979), age, and educational achievement are also factors that may influence choice of encoding strategy. The present results should be interpreted bearing in mind that the subjects were well-educated, profoundly deaf adult native signers of ASL.

The experiments reported here provide converging evidence that the distinction between item and order recall is an important one for short-term memory (Bjork & Healy, 1974; Lee & Estes, 1981; Murdock, 1976) and provide support for the hypothesis that temporal order recall may be facilitated by the use of a speech-based code (Crowder, 1978; Healy, 1975, 1977). In ordered recall tests for English letters and words (MacDougall, 1979; Pintner & Paterson, 1917; Wallace & Corballis, 1973), for fingerspelled letters (Liben & Drury, 1977) and for ASL signs (Bellugi et al., 1975), it has been found that deaf persons recall fewer items than hearing persons. The present findings are in agreement with these results. Deaf subjects in Experiment 1 responded less accurately in probed recall for order of printed English words than did hearing subjects. Furthermore, the extent to which a speech-based code was used correlated with the accuracy of ordered recall. However, in the free recall task of Experiment 2, deaf subjects did not differ significantly in recall accuracy from hearing subjects. Thus, deaf subjects seem to differ from hearing subjects in recall accuracy when recall of item and order information is required, but not when recall of only item information is required. Consistent with this hypothesis that deaf subjects may have specific difficulties with retention of temporal order information, O'Connor and Hermelin (1972, 1973) found that, given the choice of spatial or temporal order recall, deaf subjects used spatial strategies; in contrast, hearing subjects used temporal order recall strategies. Also, Lake (1980) reported that deaf children do not attend to word order when learning English.

As English is a language in which word order plays a critical syntactic role, this suggestion that deaf persons may have special trouble with recall of order information is of major interest. It is known that, on the average, deaf persons have difficulty with reading (Karchmer, Milone, & Wolk, 1979), and closer analysis shows that there are certain syntactic constructions that are particularly difficult for deaf persons to comprehend (Quigley & King, 1980). Work such as the present study on the underlying cognitive processes of deaf persons may help in understanding these reading and language problems.

REFERENCES

Baddeley, A. D. Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. *Quarterly Journal of Experimental Psychology*, 1966, *18*, 362-365.

Baddeley, A. D. The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review*, 1978, *85*, 139-152.

Bellugi, U., Klima, E. S., & Siple, P. Remembering in signs. *Cognition*, 1975, *3*, 93-125.

Belmont, J. M., & Karchmer, M. A. Deaf people's memory: There are problems testing special populations. In M. Gruneberg, P. Morris, & R. Sykes (Eds.), *Practical aspects of memory*. London: Academic Press, 1978.

Bjork, E. L., & Healy, A. F. Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 80-97.

Conlin, D., & Paivio, A. The associative learning of the deaf: The effects of word imagery and signability. *Memory & Cognition*, 1975, *3*, 335-340.

Conrad, R. An association between memory errors and errors due to acoustic masking of speech. *Nature*, 1962, *193*, 1314-1315.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, *55*, 75-84.

Conrad, R. Short-term memory processes in the deaf. *British Journal of Psychology*, 1970, *61*, 179-195.

Conrad, R. *The deaf school child*. London: Harper & Row, 1979.

Crowder, R. G. Language and memory. In J. F. Kavanagh & W. Strange (Eds.), *Speech and language in the laboratory, school, and clinic*. Cambridge, Mass.: MIT Press, 1978.

Frumkin, B., & Anisfeld, M. Semantic and surface codes in the memory of deaf children. *Cognitive Psychology*, 1977, *9*, 475-493.

Furth, H. G. Linguistic deficiency and thinking: Research with deaf subjects 1964-1969. *Psychological Bulletin*, 1971, *76*, 58-72.

Hanson, V. L. Implications of research on sign languages for theories of reading. In B. Frokjaer-Jensen (Ed.), *The sciences of deaf signing*. Copenhagen: Audiologopedic Research Group, 1980.

Healy, A. F. Coding of temporal-spatial patterns in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1975, *14*, 481-495.

Healy, A. F. Pattern coding of spatial order information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 419-437.

Hintzman, D. L. Articulatory coding in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1967, *6*, 312-316.

Karchmer, M. A., Milone, M. N., Jr., & Wolk, S. Educational significance of hearing loss at three levels of severity. *American Annals of the Deaf*, 1979, *124*, 97-109.

Kintsch, W., & Buschke, H. Homophones and synonyms in short-term memory. *Journal of Experimental Psychology*, 1969, *80*, 403-407.

Klima, E., & Bellugi, U. *The signs of language*. Cambridge, Mass.: Harvard University Press, 1979.

Koh, S. D., Vernon, M., & Bailey, W. Free-recall learning of word lists by prelingual deaf subjects. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 542-547.

Lake, D. Syntax and sequential memory in hearing impaired children. In H. N. Reynolds & C. M. Williams (Eds.), *Proceedings of the Gallaudet conference on reading in relation to deafness*. Washington, D.C.: Division of Research, Gallaudet College, 1980.

Lee, C. L., & Estes, W. K. Item and order information in short-term memory: Evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning and Memory*, 1981, *7*, 149-169.

Liben, L. S., & Drury, A. M. Short-term memory in deaf and hearing children in relation to stimulus characteristics. *Journal of Experimental Child Psychology*, 1977, *24*, 60-73.

MacDougall, J. C. The development of visual processing and short-term memory in deaf and hearing children. *American Annals of the Deaf*, 1979, *124*, 16-22.

Moulton, R. D., & Beasley, D. S. Verbal coding strategies used by hearing-impaired individuals. *Journal of Speech and Hearing Research*, 1975, *18*, 559-570.

Murdock, B. B., Jr. Item and order information in short-term serial memory. *Journal of Experimental Psychology: General*, 1976, *105*, 191-216.

O'Connor, N., & Hermelin, B. Seeing and hearing and space and time. *Perception & Psychophysics*, 1972, *11*, 46-48.

O'Connor, N., & Hermelin, B. The spatial or temporal organization of short-term memory. *Quarterly Journal of Experimental Psychology*, 1973, *25*, 335-343.

Odom, P. B., Blanton, R. L., & McIntyre, C. K. Coding medium and word recall by deaf and hearing subjects. *Journal of Speech and Hearing Research*, 1970, *13*, 54-58.

Pintner, R., & Paterson, D. G. A comparison of deaf and hearing children in visual memory for digits. *Journal of Experimental Psychology*, 1917, *2*, 76-88.

Poizner, H., Bellugi, U., & Tweney, R. Processing of formational, semantic, and iconic information in American Sign Language. *Journal of Experimental Psychology: Human Perception and Performance*, 1981, *7*, 1146-1159.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor. R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monograph*, 1969, *79*, 1-16.

Quigley, S. P., & King, C. M. Syntactic performance of hearing impaired and normal hearing individuals. *Applied Psycholinguistics*, 1980, *1*, 329-356.

Shand, M. A. Sign-based short-term coding of American Sign Language signs and printed English words by congenitally deaf signers. *Cognitive Psychology*, 1982, *14*, 1-12.

Stokoe, W. C., Jr., Casterline, D. C., & Croneberg, C. G. *A dictionary of American Sign Language*. Washington, D.C.: Gallaudet College Press, 1965.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Bureau of Publications, Teachers College, 1944.

Wallace, G., & Corballis, M. C. Short-term memory and coding strategies in the deaf. *Journal of Experimental Psychology*, 1973, *99*, 334-348.

Watkins, M. J., Watkins, O. C., & Crowder, R. G. The modality effect in free and serial recall as a function of phonological similarity. Journal of Verbal Learning and Verbal Behavior, 1974, 13, 430-447.

Wickelgren, W. A. Distinctive features and errors in short-term memory for English vowels. Journal of the Acoustical Society of America, 1965, 38, 583-588.

Wickelgren, W. A. Distinctive features and errors in short-term memory for English consonants. Journal of the Acoustical Society of America, 1966, 39, 388-398.

Appendix A

Stimulus Sets for Experiment 1

Phonetically similar set: TWO, BLUE, WHO, CHEW, SHOE, THROUGH, JEW, YOU
Phonetic control set: SOME, KING, THAT, CRY, FARM, WITH, TAX, CHURCH
Formationally similar set: KNIFE, NAME, PLUG, TENT, TRAIN, EGG, SALT, CHAIR
Formational control set: RING, COKE, RULE, MONTH, COW, HOUSE, NOON, KISS
Graphemically similar set: BEAR, MEAT, HEAD, YEAR, LEARN, PEACE, BREAK, DREAM
Graphemic control set: TREE, NORTH, GIRL, WORLD, KNOW, DRINK, WAIT, MOVE

Appendix B

Stimulus Lists for Experiment 2

Formational Similarity Condition

Set 1

Experimental list: MONTH, DURING, HAPPEN, SAME, MEET, CAN'T, DEPEND, TEMPERATURE, REGULAR, STARS, PAIN, SOCKS.

Control list: BLUE, VISIT, GROUP, READ, ACCIDENT, LAW, COMFORTABLE, WAIT, SECRET, NIECE, SOMETIMES, NEXT.

Set 2

Experimental list: NAME, RAILROAD, CHAIR, SALT, TENT, EGG, HURRY, SHORT, WEIGHT, UNIVERSE, INCREASE, VERY.

Control list: EYE, THING, GOLD, FLOWER, MARRY, UMBRELLA, BUILD, NIGHT, KEY, ABLE, HEAVEN, MEAT.

Set 3

Experimental list: STOP, TOWN, CLEAN, BECOME, PROVE, WOOD, PAPER, WINDOW, OPEN, COOK, SCHOOL, PIE.

Control list: APPLE, COW, THROUGH, PROBLEM, WARM, FAMOUS, HANDS, KING, CLEAR, TREE, ISLAND, GREEN.

Set 4

Experimental list: TEACH, NUMBER, INSIDE, BANQUET, PUT, GIVE, SMOOTH, NONE, SELL, MORE, PACK, SOIL.

Control list: DAY, SMART, BIRD, DEVIL, SUNSET, GAME, BREAD, REFUSE, COUNT, LAUGH, HOUSE, RULE.

Set 5

Experimental list: SCIENCE, COFFEE, BICYCLE, POSSIBLE, WHICH, SHOES, ADVERTISE, BREAK, HABIT, TOGETHER, MAKE, FOLLOW.

Control list: MILK, PEOPLE, TELEPHONE, RESPECT, AFTERNOON, TEASE, WATER, FIRST, SCISSORS, PRESIDENT, BEAUTIFUL, HOME.

Phonetic Similarity Condition

Set 1
Experimental list: BLUE, CHEW, TOO, THROUGH, NEW, SHOE, WHO, TRUE, FEW, TWO, YOU, KNEW.

Control list: SICK, PACK, ALL, BREATHE, RED, TIME, COP, MORE, HOT, OUT, BOY, PLAY.

Set 2
Experimental list: WEIGH, GREAT, PRAY, SKATE, EIGHT, THEY, LATE, DAY, STRAIGHT, ATE, WAIT, GRAY.

Control list: SMELL, RIGHT, HUNT, SNAKE, LARGE, THAT, RICH, ICE, STRENGTH, AID, PLAY, BALD.

Set 3
Experimental list: FREEZE, PIECE, PLEASE, THESE, PEAS, EAST, TEASE, CHEESE, GREECE, PEACE, NIECE, PRIEST.

Control list: PREACH, PLANT, PRAISE, THEIR, LUCK, HERE, SPELL, THRILL, PURSE, TRAIN, CLOWN, THIEF.

Set 4
Experimental list: CALM, DAWN, FROM, BOMB, ONE, SOME, GONE, FUN, DONE, COME, MOM, THUMB.

Control list: NEED, LIST, ELSE, PLUS, JOY, REAL, BORN, CAT, FINE, POOR, ART, MOUSE.

Set 5
Experimental list: TRY, LIE, EYE, FLY, PIE, WHY, DIE, GUY, MY, HIGH, BYE, DRY.

Control list: CRY, END, LAW, GET, PEN, MAD, EAT, OWL, WE, LONG, JOG, OLD.

A COMMON BASIS FOR AUDITORY SENSORY STORAGE IN PERCEPTION AND IMMEDIATE MEMORY*

Robert G. Crowder+

Abstract. Thirty-two subjects participated in three experiments, one assessing auditory short-term memory for word lists with and without a verbal suffix and two assessing discrimination of synthetic vowels at either short or long interstimulus delays. The purpose was to find out whether the same kind of auditory memory supports both short-term memory and speech discrimination. There was a significant correlation between performance in the suffix and A-X speech-discrimination experiments in those conditions likely to depend partly on echoic memory; however, there was no significant correlation between the tasks in conditions in which echoic memory was presumed to have been removed. The results provide a bridge between perception and memory procedures and support a theoretical model that was made to cover both domains.

The suffix effect is a decrement in recall of the last item in an immediate-memory list caused by an extra utterance (which does not have to be recalled) presented at the end of the list. Since the paper by Crowder and Morton (1969), one influential hypothesis for this phenomenon has been that a verbal suffix damages information that otherwise remains available, in sensory form, following auditory presentation. A survey of the research supporting that general position is available in Crowder (1976, Chapter 3) and a recent, specific version of the hypothesis is in Crowder (1978). The hypothesis is that speech sounds are represented, after they occur, on a two-dimensional, neurally spatial grid that is organized by input channel and time of arrival. The entries on this grid are spectral descriptions of the speech sounds, similar to sound spectrograms. It is assumed (Crowder, 1978) that these representations are related to each other through the rules of recurrent lateral inhibition. From this, it follows that after a series of utterances on the same physical channel (i.e., the same voice in the same location), there will be lingering auditory information about the most recent arrival.

---

*Also in Perception & Psychophysics, 1982, 31(5), 477-483.

+Also Department of Psychology, Yale University, New Haven, CT.

Acknowledgment. This research was supported by NSF Grant BNS 8005838 to R. Crowder and by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories. I acknowledge with pleasure the assistance of Virginia Walters for her help in all phases of this research, of Danny Kaye for warning me about the perils of correlating difference scores, and of Robert Sternberg and Bruno Repp for commenting on an earlier version of this paper.

[HASKINS LABORATORIES: Status Report on Speech Research SR-70 (1982)]

This most recent item will be receiving lateral inhibition from only one direction, as opposed to the earlier items, which are inhibited from two directions. (The first few items in the series, including the very first, would not be prominent in the auditory system because of the sheer amount of time they have been undergoing mutual inhibition.) The freedom of the last utterance in a series from retroactive lateral inhibition is held responsible for the large recency effect observed in immediate memory tests with auditory presentation, but not with visual presentation (which does not activate the system under consideration here).

When a redundant suffix item is presented on the same channel as the memory list, just following the last to-be-remembered item, the latter loses its special status of being free from lateral inhibition from one direction, causing the suffix effect. The availability of this residual information about how the most recent item sounded is presumably used by the subject to supplement his regular categorical short-term memory for the items. This regular short-term memory is roughly the same whether the input modality is visual or auditory, but the auditory residual about the most recent item gives the latter modality the edge when the two are compared.

There are several recent pieces of research that may well force significant revision of this hypothesis for auditory memory (Ayres, Jonides, Reitman, Egan, & Howard, 1979; Campbell & Dodd, 1980; Spoehr & Corin, 1978); however, the form of such a revision will likely leave intact the major assumptions about the suffix and modality effects and their common dependence on the same system (e.g., see Morton, Marcus, & Ottley, 1981). It is probably fair to say that competing interpretations of the suffix effect have not yet been so thoroughly worked out as the one offered above. For example, those that propose specific hypotheses about how the suffix works often leave unexplained the modality effect (Spoehr & Corin, 1978). Other competitors, such as the attention-grouping suggestions of Kahneman and Henik (1981), seem to be dealing with a less molecular level of analysis than the explanation outlined above. When "grouping," for example, is used to explain something, the next question is always, "What causes grouping?" Indeed, an explanation of grouping in the auditory system might well rely on principles of lateral inhibition!

In the speech-perception literature, it has been explicitly claimed for years that auditory memory plays an important role in speech discrimination experiments (Pisoni, 1973, 1975; Pisoni & Tash, 1974; Fujisaki & Kawashima, Note 1). The original idea here was that if phonetic category differences are not available to discriminate two similar speech tokens, they must be discriminated on the basis of their sounds. Since the sounds to be distinguished cannot ordinarily be presented simultaneously, this requires that the earlier item be remembered in sensory form until the later item, with which it is to be compared, has arrived.

The process assumption was that subjects try first to discriminate speech sounds on a phonetic basis and then go on to consult auditory memory only if the phonetic test fails. This "phonetic first" dual coding hypothesis has not fared very well empirically (Crowder, 1982; Pisoni, 1973; Repp, Healy, & Crowder, 1979). These studies all varied the delay between two vowels being discriminated in the A-X (same/different) paradigm. It would be expected that, according to the dual-coding hypothesis, the within-category discriminations would depend more on auditory short-term memory than the between-category discriminations. On the reasonable assumption that auditory memory decays faster than phonetic memory, then, the effect of a delay between the items being discriminated should be larger for the within- than for the between-category trials. Although Pisoni (1973, p. 258) reported this outcome verbally, there was a ceiling effect on between-category performance in discrimination hits (calling a true DIFFERENT trial "different"), and, with the d' performance measure, the decay slopes for within- and between-category trials were parallel. Crowder (1982) and Repp et al. (1979) obtained just the same result, parallel decay for within- and between-category discriminations along vowel continua, as a function of interitem delay.

However, the case for some role of auditory memory in vowel discrimination is a rather strong one, even if the phonetic-first, dual-code hypothesis is wrong; the fact that interstimulus delay causes deterioration in A-X vowel discrimination, by itself, is supportive of some role for auditory sensory memory in the task. This occurred reliably in the Pisoni (1973), Repp et al. (1979), and Crowder (1982) experiments. Furthermore, Pisoni (1975) showed that an interpolated vowel sound, placed immediately after target tokens in the ABX paradigm, significantly reduced performance compared with white-noise and tone controls. Repp et al. (1979) replicated this interference effect in the simpler, A-X task, by placing the iterference sound midway between the two items being discriminated. Repp et al. suggested that this interference effect was the same disruption of auditory sensory memory that is observed in the suffix experiment.

The present experiments are aimed at strengthening the argument that the same auditory memory system serves both the suffix and vowel-discrimination tasks. The approach to be used relies on analysis of individual differences, rather than on experimental comparisons. The experimental work done in the past has produced three lines of evidence for a common auditory memory system in perception and short-term memory. The first point is the interference mentioned just above: in both the memory and perception experiments, an extra utterance seems to prevent the use of sound information for what just preceded the interfering item. In the suffix situation, it is the suffix that masks auditory memory for the last item on the list. In the vowel-discrimination setting, the masking vowel comes between the two sounds being distinguished in the A-X task (Repp et al., 1979).

The second point is that auditory memory in both situations seems to be subject to temporal decay. Crowder and Morton (1969) suggested that a life of approximately 2 sec would be a plausible figure for the suffix experiment, and I have recently demonstrated (Crowder, 1982) that vowel-discrimination performance reaches asymptote when the A-X delay interval is approximately 3 sec.

The third point of similarity between the suffix and vowel-discrimination tasks is their common dependence on the phonetic class involved in the experiment. Pisoni (1973) first showed that the decay in A-X discrimination was much greater for stop consonants than for steady-state vowels. Crowder (1971) demonstrated that neither the modality effect nor the suffix effect occurs when the lists to be remembered contain items distinguished only by initial stop consonants. Crowder (1973) also demonstrated the same result with terminal stops. The fact that presumptive auditory-memory contributions come and go together as a function of phonetic class, in the two experimental settings, is consistent with the idea that they represent two manifestations of a common memory system.

Another strong point favoring this interpretation would be if individual subjects who showed a large auditory-memory capacity in the suffix task also showed a large auditory-memory capacity in the discrimination task. This outcome would cement the case for a common processing system in the two settings. But there are at least two circumstances that are discouraging from the very start of such an investigation of individual differences. One is that the auditory-memory contribution is numerically a small one compared with the effects of other variables in both experiments. The suffix effect is robust, but it is small in magnitude compared with the inventory of other established processes in immediate memory (encoding common to visual and auditory input, grouping, rehearsal, etc.). In vowel discrimination as well, the portion of performance that is sensitive to A-X delay, and therefore presumably the portion that shows auditory memory, is numerically very small (Crowder, 1982). So there is the risk that the performance components of interest are inherently swamped by other factors in any real experimental setting.

The second cautionary note is that the type of memory under consideration here may simply not differ much among people. If auditory memory in these settings is truly as sensory as has been claimed (Crowder, 1978), one might expect it to be relatively invariant and uninteresting from an individual-differences standpoint. This is not to say that people are equivalent in their sensory capacities, of course. Indeed, it is hard to know how one could ever establish that people differ more in, say, working memory capacity than they do in visual acuity. However, in the context of tasks that are weighted more toward the complicated than toward the simple cognitive functions, it must be considered risky to be searching for individual differences in the simpler components. (An extreme example would be looking for individual differences based on visual acuity in the context of visually presented analogy problems.) For all these reasons, a negative outcome would not eliminate the case for a common memory system, but a positive outcome would be a striking victory for the theory.

METHOD

The subjects were taken through one suffix experiment and two vowel-discrimination experiments. The suffix effect has been well behaved in our laboratory for some time, and therefore there was little question how to conduct that part of the investigation. However, there are a number of possible discrimination paradigms, and it seemed undesirable to rely on only a single one of these. The traditional paradigm of choice in speech perception was for many years the so-called ABX paradigm, in which people hear three tokens of which the first two are different and they must decide whether the third is equal to one or the other of these first two. It has been claimed more recently (e.g., see Best, Morrongiello, & Robson, 1981) that the ABX procedure systematically discounts auditory memory. This is because the second item in the ABX triad could serve to mask auditory storage of the first item until the third one arrives, and subjects may adopt the strategy of trying to compare the trace of the second item with the third. The A-X (same-different) procedure would seem better suited for showing auditory-memory effects because nothing comes between the two items being distinguished. By collecting data on the same stimuli and the same subjects in both ABX and A-X procedures, it would be possible to compare the reliabilities and sensitivities of the two procedures directly. However, the main reason for using both ABX and A-X procedures was not to compare their sensitivities formally (which would require a much more extensive experiment to be definitive) but, rather, to maximize the chances of getting at least one discrimination task that could be associated with short-term memory.

The Suffix Experiment

Subjects and materials. The subjects were 32 young adults of both sexes from our summer subject pool. Most, but not necessarily all, of them were college students during the academic year and were paid for their participation.

The stimuli were the nine digits, the nonsense syllable "ba," and a 1,000-Hz tone. The verbal items were recorded by a male speaker and digitized on the Haskins Laboratories Pulse Code Modulation system, each in a 450-msec time slot. These items were then accessible independently to other computer routines for automatically assembling the actual stimulus lists.

Design and procedure. There were 20 trials in which nine-digit series were followed by the 1,000-Hz tone and 20 in which the series were followed by the verbal suffix "ba." On each trial, there was a 250-msec pause between each of the digits and between the last memory item and the redundant suffix or tone. Subjects were allowed 20 sec for ordered written recall after each trial. Since there was no interest in looking at subtle poperties of the suffix effect here, all subjects received the 20 control (tone) trials first and the 20 suffix trials second. (It will be seen below that not counterbalancing order of stimuli had no apparent effect on the suffix experiment as compared with numerous data sets in the literature in which these precautions were followed.) The instructions were standard in that they emphasized ordered recall and characterized the extra item (suffix or tone) as a cue telling people when to begin their recall attempt.

The Discrimination Experiments

The ABX and A-X experiments were conducted on the same 32 subjects as in the suffix experiment and directly after it. These two discrimination procedures were used in counterbalanced order, half the subjects starting with one and half with the other.

Stimuli. The stimulus items were all 300-msec steady-state synthetic vowels produced on the Haskins Laboratories OVE IIIc synthesizer. There were eight different tokens ranging from /i/ to /I/ in approximately equal steps. The fundamental frequency for all tokens was brought from 90 to 100 Hz during the first 100 msec, remained at 100 Hz for the interior 100 msec, and then dropped to 85 Hz during the final 100 msec. The eight center frequencies of the first, second, and third formants, respectively, were: F_1--269, 287, 304, 320, 339, 356, 372, and 391; F_2--2,198, 2,167, 2,136, 2,105, 2,075, 2,045, 2,016, and 1,987; F_3--3,019, 2,933, 2,870, 2,809, 2,749, 2,690, 2,613, and 2,557. Overall amplitude for the vowels was constant over their duration. The materials were presented over loudspeakers at a comfortable level in a relatively quiet room.

Design. There were four blocks of speech-discrimination trials. For half of the subjects, the first two were ABX, and the second two were A-X; for the other half, this was reversed. The test stimulus (X) for either kind of discrimination trial was spaced at either a short (500-msec) or a long (3,000-msec) delay relative to the comparison stimulus (A in A-X or B in ABX tests). This was to affect the presence of auditory memory; details are given in the following sections. The design feature common to both discrimination procedures was that, for each task (ABX and A-X), half the subjects had the short interval first and half had the short interval second. In other words, the scheduling of delay intervals across the four blocks of discrimination trials was either short-long-short-long or it was long-short-long-short. Again, there seemed no reason to avoid confounding the short-long order in the two paradigms because the project was aimed at individual differences rather than at point estimates for experimental effects.

The ABX task. On each ABX trial, there was first a 1,000-Hz tone, followed, after a 250-msec delay, by the first of three vowel tokens relevant to that trial. Then, following a delay that was always set at 250 msec, the second of the three vowel sounds occurred. These first two vowels were always different tokens from the eight-item continuum. The delay between the second and the third of the items was the one that was varied to affect auditory memory decay; it was either 500 or 3,000 msec. There was then a 2,000-msec delay for the subject to record his response.

All possible one-step and two-step discriminations were tested in the ABX task. Consider the first two of the three vowels presented on a trial and call the eight vowels 1,2,...,8. There are 14 one-step combinations (1-2, 2-1, 2-3, 3-2, 3-4, 4-3, etc.), and each of these has to be presented twice so that the correct answer is equally often the choice of "A" and "B" in the ABX triple (1-2-1, 1-2-2, 2-1-1, 2-1-2, 2-3-2, 2-3-3, etc.). Thus, there must be 28 different one-step trials. Analogously, there are 24 different two-step trials (1-3-1, 1-3-3, 3-1-3, 3-1-1, 2-4-2, 2-4-4, etc.). The 52 possible ABX trials were each presented once in the short-delay version and once in the long-delay version, for a total of 104 ABX trials per subject. Within these constraints, the order of trials was random.

The ABX instructions stated that the first two vowels in a triple would always be different and that subjects should circle the number "1" or "2" on the answer sheet, depending on which of the first two vowels they thought matched the third.

The A-X task. The A-X task routine is, of course, simpler than the ABX because there are only two events on each trial instead of three. Following the tone, there was a 250-msec pause, which was then followed by the first of the two vowels to be discriminated. After either a 500- or a 3,000-msec delay, the second vowel occurred, and the subject had 2,000 msec to make his or her same-different response before the next trial started. The same 52 stimulus pairs used in ABX testing--28 one-step and 24 two-step--were presented as the "different" trials in A-X. However, an additional 16 "same" pairs were added in which the two vowels were physically identical (1-1, 1-1, 2-2, 2-2, etc.). This meant that a complete replication contained 68 trials, and two such replications were carried out, one for the short delay and one for the long delay. Instructions for the A-X procedure simply asked the subjects to circle the letters "s" or "d" on each trial, depending on whether or not the two vowels seemed to be "exactly the same sound."

RESULTS

The results will be presented in several sections. First, it will be established that each of the three separate experiments in this set produced reasonable results on its own, in terms of the existing literature. This is very much a precondition for examining individual differences among them. Second, the issue of formal reliability will be raised for the three data sets; this is another precondition, for if the measures are not reliable, there will be little use looking for individual differences. Finally, correlations among the different tasks will be considered.

The Suffix Experiment

Figure 1 shows the basic result of the suffix experiment. Every one of the subjects showed more errors in the suffix condition than in the control condition. For each condition there were 180 possible errors (20 trials X 9 positions); the mean errors for the control and suffix conditions were, respectively, 42.75 and 69.74 [$t(30) = 8.19$, $p < .0005$]. It is clear from the figure that the difference was located mainly toward the end of the list, most especially at the last serial position. In relation to the published literature, then, this was a thoroughly routine suffix experiment.

The Discrimination Experiments

Table 1 shows summary statistics from the ABX and A-X discrimination procedures. If these procedures are good tests of discrimination, it is reasonable to expect a large effect of step size, which in these experiments was set at either one or two. (In ABX tests of the continuum 1,2,...,8, a one-step trial might present 1-2-2 and a two-step trial might present 1-3-3; in A-X tests, the corresponding trials could be 1-2 and 1-3.) The first section of Table 1 shows that indeed both procedures led to markedly fewer errors for the two-step than for the one-step trials. The ABX procedure, however, gave a smaller value of t than the A-X procedure, 11.91 vs. 23.39.

The lower half of the table shows the data split according to the length of the delay interval, either the interval between A and X in the A-X task or the interval between B and X in the ABX task. In both discrimination procedures, there was a higher error rate when this interval was long than when it was short; however, the difference was statistically significant only in the A-X task.

The data on discrimination as a function of delay were further examined using the tables of Kaplan, Macmillan, and Creelman (1978) for calculating d'

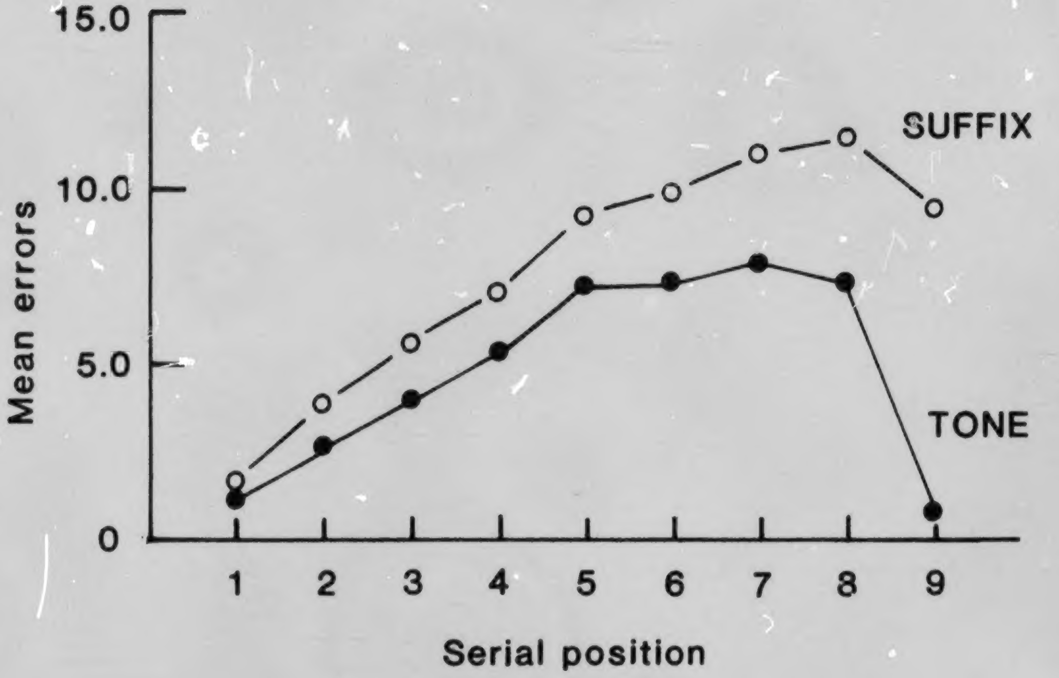


Figure 1. The relation between errors and input serial position in the suffix experiment.

Table 1

Speech Discrimination: Summary Statistics for Error Proportions

		Task	
Comparison		ABX	A-X
Step Size	One Step	.401	.587
	Two Step	.207	.277
	t(30)	11.914	23.391
Delay	Short	.306	.344
	Long	.317	.445
	t(30)	.516	8.223

Table 2

Sensitivity (d') as a Function of Task and Delay

	Task	
Interval	ABX	A-X
Short	2.801	3.501
Long	2.247	2.910
t(7)	4.414	4.547

Table 3

Odd-Even Reliabilities

Measure	Coefficient Raw Correlation	 Spearman-Brown Corrected
1. Suffix Experiment		
a. Total Suffix Errors	.937	.967
b. Total Control Errors	.927	.962
2. Discrimination Experiments		
a. Total ABX Errors	.315	.479
b. Total A-X Errors	.667	.800
3. A-X Discrimination		
a. Short Delay	.631	.774
b. Long Delay	.470	.639

Note: All correlations in left column reliable at $p < .01$ except 2.a., which is not significantly different from zero, $t(30) = 1.818$.

from different discrimination paradigms. This analysis is shown in Table 2, in which the data are averaged over eight "supersubjects" of four individuals each, a grouping that was intended to minimize hit and false-alarm rates approaching zero and unity. The four subjects within a supersubject shared exactly the same counterbalancing condition: There were two such control variables--whether ABX preceded A-X, or the other way around, and whether the short interstimulus intervals were tested first or second within each paradigm. Thus, there were four possible arrangements, and eight subjects, making up two supersubjects, received each. If Kaplan et al. (1978) are correct in asserting that these are fair measures of sensitivity across paradigms, then it may be concluded that the A-X task gives better discrimination than the ABX task [$t(7) = 6.37$, $p < .0005$]. However, by this measure, the delay effect was reliable for both paradigms.

These analyses indicate that both discrimination experiments produced plausible results but that the A-X procedure might be more sensitive and therefore more useful for analyzing individual differences. The same conclusion comes from a formal analysis of reliability, which comes next.

Reliability

The best measure of the suffix effect, for purposes of ordinary experimentation, is probably some difference score, or ratio, representing how recency is changed on the last position across the suffix and control conditions. Although such measures have been useful for at least a decade of experimental work, they turn out to have limited reliability in individual differences analysis. Several such "pure" measures of the suffix effect, which show the group data of Figure 1 to good effect, gave odd-even reliabilities that were not significantly different from zero. The unreliability of difference scores is well documented (Cronbach & Furby, 1970; Guilford, 1956).

The strategy followed here was to concentrate on measures from the suffix experiment that included, according to the theory, or did not include, according to the theory, a contribution from auditory memory. The control condition should contain this contribution, and the suffix condition should not. Table 3 shows the odd-even reliabilities of the total number of errors made in the control and suffix conditions, with and without the Spearman-Brown correction for attenuation. The odd-numbered trials were simply correlated with the even-numbered trials, over subjects, to produce these reliabilities. The Spearman-Brown correction enters the picture because there are only half as many observations in the two halves being correlated as there were on the original test. These reliabilities are highly reassuring and suggest that one could have designed this project with a shorter period of testing in the suffix experiment.

The odd-even reliabilities of the total errors made in the ABX and A-X situations are also entered in Table 3, with and without the Spearman-Brown correction. (As in the suffix experiment, scores based on differences between the short and long delay interval--which should, theoretically, have been purer measures of auditory memory--were not at all reliable.) There is a clear basis for distinguishing the reliabilities of the ABX and A-X procedures here. The A-X procedure is more than twice as reliable, in the uncorrected data, as the ABX procedure. This may or may not be a general result: It is at least consistent with the stronger statistical evidence for step-size effects and for delay effects found in A-X compared with ABX testing. To repeat what was said earlier, the main purpose of this comparison was to come up with a suitable measure for comparing the suffix and discrimination experiments, not choosing the "best" discrimination task. Nonetheless, this result does suggest some caution for investigators choosing the ABX task, lest they be making it hard for themselves to demonstrate experimental effects in a sensitive way.

The third section of Table 3 shows odd-even reliabilities for the two main conditions of A-X discrimination, the short and long conditions. These ought to represent A-X discrimination with and without, respectively, the benefit of auditory memory, or, at least, there ought to be more auditory memory in the short than in the long condition. These reliabilities are satisfactory, although not as impressive as those that came from the suffix experiment.

The Relation Between Immediate Memory and Discrimination

From the suffix experiment and from the A-X discrimination experiment, there are two scores for every subject, one in each experiment likely to include performance based on auditory memory and another likely not to include auditory memory. In the suffix experiment, the total performance in the control condition would be expected to include auditory memory but not performance in the suffix condition, because the suffix would have removed that component. In the A-X experiment, there should be an auditory component at the short interstimulus interval but not at the long interval, at which the auditory trace would have decayed.

Table 4 shows the relevant correlations. Notice, first, that there are large correlations between the two measures from both of the tasks. This indicates that there is a great deal of shared variance within either the suffix or discrimination experiments that, presumably, has nothing to do with auditory memory. In the upper right-hand quadrant of the table, the correlations are quite a bit lower, representing the relation between memory and speech discrimination. Of these four correlations, the only one that is different from zero, statistically, is the one that is presumed to contain the common component deriving from auditory memory. This reliable correlation of .367 ($p < .025$) is the major positive result of this set of experiments. In psychometric terms, it is not impressive in size, representing shared variance of about 13.5% between the two tasks. However, these psychometric criteria are not usually applied to data from straight experimental designs, for some reason. In terms of experimental work, rather, investigators typically celebrate when an a priori prediction specifying one of four conditions to exceed the other three comes out at better than the .025 level of confidence.

Table 4

Correlations Within and Between Memory and Discrimination Tasks

	Total Control Errors	A-X Short Errors	A-X Long Errors
Total Suffix Errors	.853	.278	.262
Total Control Errors		.367	.272
A-X Short Errors			.731

Note: t(30) values for .278, .262, .367, and .272, respectively, are 1.59, 1.49, 2.16, and 1.58.

The highest of the other between-task correlations was .278 ($p > .05$). These other, nonsignificant, intertask correlations show that it was not just some general factor such as motivation or intelligence that produced the target relationship, for those factors would have led to relationships between all measures from the two experimental tasks. Rather, it must be counted a victory for the theory that the significant relationship occurred precisely where it was supposed to and nowhere else. (This is not to imply a much larger number of subjects would not push the three other intertask correlations to statistical reliability. There are other factors that might produce common variance in different laboratory tasks. The main point is that, within this particular study, it was only the expected correlation that was reliable.)

Furthermore, the obtained correlation of .367 is not quite as meager as it first seems. The square root of the reliability coefficient sets an upper limit on the variance that can be accounted for when the mesuare is correlated with anything external (validity). The square root of the odd-even reliability of total errors in the A-X short condition is .817. The variance in the total errors from the control condition in the suffix experiment, accounted for by A-X short errors, was .135. Thus, the discrimination measure accounted for about 16.5% (.135/.817) of the reliable variance in the suffix measure, which is not a disgrace considering the huge number of other components in both tasks.

DISCUSSION

One form of explanation in psychology is to relate the known properties of an experimental procedure to concepts that are more general than that specific procedure. It is often not terribly hard to offer a model for an experiment like the suffix experiment that accommodates its various properties neatly. Still, if the components of that model have no generality outside the suffix experiment, we are not satisfied that a true explanation has occurred. It is necessary to generalize components of the model to other settings in order to have a satisfying explanation.

There are several ways to establish generality of components across tasks. One is to show that the same experimental variables influence performance in the same way in each of two tasks. This much has been done in several areas. In short-term-memory experiments, for example, it has been shown that the suffix effect and also the visual-auditory modality effect disappear when the memory stimuli are distinguished only by stop consonants. Pisoni (1973) showed the vowel-stop consonant difference in speech discrimination. Likewise, interpolating an unrelated masking sound has a comparable interfering effect in both the memory and vowel-discrimination experiments. Thus, the two task settings respond quite similarly to certain experimental manipulations.

A second means of generalizing concepts across task settings is represented in this work--showing that individual differences in a theoretically specific component correlate reliably across the two tasks. People who show outstanding auditory memory in the immediate-memory control condition also show outstanding auditory memory in the A-X task with a short interstimulus interval. No single approach to this generalization of concepts is sufficient by itself, but when they operate in parallel, as they seem to here, one is justified in placing more weight on the explanatory power of the model in question. In this case, there seems to be even more reason, then, to take seriously the possibility that speech perception and short-term memory have some important information-processing processes in common.

REFERENCE NOTE

1. Fujisaki, H., & Kawashima, T. *On the modes and mechanisms of speech perception* (Annual Report of the Engineering Research Institute). Tokyo: University of Tokyo, 1969.

REFERENCES

Ayres, T., Jonides, J., Reitman, J. S., Egan, J. C., & Howard, D. A. Differing suffix effects for the same physical suffix. *Journal of Experimental Psychology: Human Learning and Memory*, 1979, *5*, 315-321.

Best, C. T., Morrongiello, B., & Robson, R. Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, 1981, *29*, 191-211.

Campbell, R., & Dodd, B. Hearing by eye. *Quarterly Journal of Experimental Psychology*, 1980, *32*, 85-89.

Cronbach, L. J., & Furby, L. How we should measure "change"--or should we? *Psychological Bulletin*, 1970, *74*, 68-80.

Crowder, R. G. The sounds of vowels and consonants in immediate memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 587-596.

Crowder, R. G. Representation of speech sounds in precategorical acoustic storage. *Journal of Experimental Psychology*, 1973, *98*, 14-24.

Crowder, R. G. *Principles of learning and memory*. Hillsdale, N.J.: Erlbaum, 1976.

Crowder, R. G. Mechanisms of backward masking in the stimulus suffix effect. *Psychological Review*, 1978, *85*, 502-524.

Crowder, R. G. Decay of auditory information in vowel discrimination. *Journal of Experimental Psychology: Human Learning, Memory, and Cognition*, 1982, *8*, 153-162.

Crowder, R. G., & Morton, J. Prcategorical acoustic storage (PAS). *Perception & Psychophysics*, 1969, *5*, 365-373.

Guilford, J. P. *Fundamental statistics in psychology and education*. New York: McGraw-Hill, 1956.

Kahneman, D., & Henik, A. Perceptual organization and attention. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, N.J.: Erlbaum, 1981.

Kaplan, H. L., Macmillan, N. A., & Creelman, C. D. Tables of d' for variable-standard discrimination procedures. *Behavioral Research Methods & Instrumentation*, 1978, *10*, 796-813.

Morton, J., Marcus, S. M., & Ottley, P. The acoustic correlate of "speech-like": A use of the suffix effect. *Journal of Experimental Psychology: General*, 1981, *110*, 568-593.

Pisoni, D. B. Auditory and phonetic codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 1973, *13*, 253-260.

Pisoni, D. B. Auditory short-term memory and vowel perception. *Memory & Cognition*, 1975, *3*, 7-18.

Pisoni, D. B., & Tash, J. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 1974, *15*, 285-290.

Repp, B. H., Healy, A. F., & Crowder, R. G. Categories and context in the perception of isolated, steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, *5*, 129-145.

Spoehr, K. T., & Corin, W. J. The stimulus suffix effect as a memory coding phenomenon. *Memory & Cognition*, 1978, *6*, 583-589.

PHONOLOGICAL AWARENESS AND VERBAL SHORT-TERM MEMORY: CAN THEY PRESAGE EARLY READING PROBLEMS?

Virginia A. Mann+ and Isabelle Y. Liberman++

Abstract. Many studies have established an association between early reading problems and deficiencies in certain spoken language skills, such as the ability to become aware of the syllabic structure of spoken words, and the ability to retain a string of words in verbal short-term memory. A longitudinal study now shows that inferior performance in kindergarten tests of these same skills may presage future reading problems in the first grade. Based on these findings, procedures are suggested for kindergarten screening and for some ways of aiding children who, by virtue of inferior performance on these tests, might be considered at risk for reading failure.

The deficiencies of poor beginning readers in certain language skills have now been amply documented. As compared to successful beginning readers, for example, these children tend to be less aware of the phonological structure of spoken words (Fox & Routh, 1975; Golinkoff, 1978; Liberman, Shankweiler, Fischer, & Carter, 1974; Rosner & Simon, 1971). They may also fall behind good readers in their short-term memory for such linguistic material as a string of letters (Liberman, Shankweiler, Liberman, Fowler, & Fischer, 1977; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979), a string of words (Mann, Liberman, & Shankweiler, 1980), or even the words of a sentence (Mann et al., 1980; Wiig & Semel, 1976).

In previous work, our concern has been the association between deficiencies in these skills and reading disability in the elementary grades. Now we turn to the question of whether a deficiency in either skill not only characterizes disabled readers in the primary grades but may indeed be found to be an early sign of reading problems. More specifically, we ask whether

---

+Also Bryn Mawr College.
++Also University of Connecticut.

Acknowledgment. This work was partially supported by a grant from the National Institute of Child Health and Human Development (Grant HD-01994) to Haskins Laboratories and by a NICHD postdoctoral fellowship (Fellowship HD05677) to Virginia A. Mann. Thanks are due to Alvin Liberman and Donald Shankweiler for their comments on earlier versions of this manuscript. We wish to express our appreciation for the generous cooperation of the staff of the Parker Memorial School in Tolland, Connecticut. Special thanks are due the principal, Robert Lincoln, and the kindergarten and first-grade teachers. We are also grateful to the children who participated and to the parents, without whose permission we could not have proceeded.

[HASKINS LABORATORIES: Status Report on Speech Research SR-70 (1982)]

reading problems in the first grade may be signalled by deficient language skills in kindergarten. We ask this question out of a consideration of the role that each skill might play in the process of reading acquisition. First, it seems likely to us that an awareness of the phonological structure of speech is necessary if one is to "crack the code" of an alphabetic system. As we have noted previously (Liberman, 1971, 1973; Liberman, Liberman, Mattingly, & Shankweiler, 1980; Liberman & Mann, 1981), the alphabet does represent the phonological structure of words more or less accurately, and a child who is unaware of that structure must be at a serious disadvantage in reading new words. Second, it seems obvious to us that the comprehension of a sentence, whether written or spoken, requires the short-term retention of many of the component words of that sentence. Therefore, we would expect that the processing of either spoken or written language would demand an ability to store verbal material efficiently in short-term memory (Liberman, Mattingly, & Turvey, 1972).

Considerable indirect evidence from widely diverse subject populations shows that a strong positive relation exists between children's awareness of the phonemic and syllabic structure of speech and their success in learning to read (Fox & Routh, 1975; Golinkoff, 1978; Liberman et al., 1974; Rosner & Simon, 1971). There is even some evidence that a deficiency in phonological awareness in a kindergartener may presage problems in beginning reading (Goldstein, 1976; Liberman et al., 1974). Less is understood, however, about the relation between early reading proficiency and short-term memory for verbal material. Moreover, even less is known about whether awareness of phonological structure and verbal short-term memory skill are correlated. On the one hand, it seems entirely possible that deficiencies in these two abilities may be relatively independent. It is also possible, however, that an adequate means of storing an utterance in short-term memory is necessary if one is to manipulate the syllabic or phonemic structure of that utterance. It is even conceivable that conscious awareness of phonological structure may somehow facilitate the use of phonetic representation in short-term memory.

In an attempt to clarify the interrelationships among phonological awareness, verbal short-term memory, and beginning reading ability, we have conducted a two-year longitudinal study, in which we tested children first as kindergarteners and subsequently as first graders. As kindergarteners, each of our subjects received a series of four different tests: a test of phonological awareness, a test of verbal short-term memory, a test of nonverbal short-term memory, and a test of IQ. As first graders, they again received the verbal and nonverbal short-term memory tests, and were, in addition, given a test of reading ability.

As our test of phonological awareness, we chose a syllable counting test (Liberman et al., 1974). In that test, children "tap out" the number of syllables in spoken words such as "bag" and "butterfly." Performance on this test has been found to be a fairly adequate predictor of reading success in the first grade, if not quite so successful as the analagous phoneme counting test (Liberman et al., 1974). We chose to test syllable segmentation rather than phoneme segmentation because syllable segmentation ability is not easily confounded by reading instruction, whereas phoneme segmentation may to some degree be reciprocally related to reading skill (Alegria, Pignot, & Morais, in press; Morais, Carey, Alegria, & Bertelson, 1979). That is, whereas phoneme segmentation ability may be helpful in the development of reading skill, increased reading skill may itself also accelerate development of phoneme awareness.

The materials used for testing children's verbal short-term memory skill were four-item word strings designed along the lines of those used in Mann et al. (1980). That study had involved a procedure in which children's performance in recalling strings of phonetically confusable (rhyming) words is compared with that for strings of phonetically nonconfusable (nonrhyming) words. Whereas the phonetically nonconfusable words allow subjects to make optimal use of the mature strategy of using phonetic representation as a means of retaining verbal material in short-term memory, the phonetically confusable words penalize the use of phonetic representation (Baddeley, 1978; Conrad, 1964). Thus the difference between performance on the two types of word strings may provide an index of the extent to which subjects rely on phonetic representation in short-term memory. Our past results reveal that good beginning readers typically surpass poor beginning readers in recall of phonetically nonconfusable word strings, but at the same time are more penalized by the manipulation of phonetic confusability. We have interpreted this finding as evidence that the inferior recall of poor readers may be due to an inability to make effective use of phonetic representation in working memory—a conclusion that we first offered to account for findings obtained in a study of letter string recall (Liberman et al., 1977; Shankweiler et al., 1979) and subsequently extended to findings obtained in a study of word string and sentence recall (Mann et al., 1980). Our question in the present longitudinal study is whether, among kindergarteners, a relatively poor memory for word strings, coupled with a relative tolerance for the effects of phonetic confusability, will presage reading difficulty in the first grade.

Elsewhere (Katz, Shankweiler, & Liberman, 1981; Liberman, Mann, Shankweiler, & Werfelman, in press), we have argued that the short-term memory difficulties of poor beginning readers are limited to the domain of verbal memory (perhaps as a specific consequence of a problem with the use of phonetic representation). Consistent with this view, there is evidence that though good and poor readers differ in verbal short-term memory, they are equivalent in recall of nonverbal material such as "doodle" designs (Katz et al., 1981; Liberman et al., in press) and photographs of unfamiliar faces (Liberman et al., in press). The present study afforded us an opportunity to gain further evidence pertinent to this issue. To that end, we included a nonverbal short-term memory test, the Corsi block test (Corsi, 1972), in our test battery. That test, which requires subjects to recall sequentially presented visuospatial information, has been used successfully in differentiating patients with lesions of the right and left hemispheres. Whereas verbal short-term memory performance has been found to suffer as a consequence of damage to the left or language-dominant hemisphere, memory performance on the Corsi blocks is impaired by damage to the right or nondominant hemisphere (Corsi, 1972; Milner, 1972).

METHOD

Subjects

The subjects in this study attended the public schools in Tolland, Connecticut. Each of them was first seen during May of kindergarten and again during May of first grade. Of the initial subject pool, which consisted of all pupils in each of four kindergarten classes, only eight children were not available for subsequent testing as first graders. The final population consisted of 62 children, 31 girls and 31 boys, whose mean age at the time of the first experimental session was 70.3 months.

Materials

As kindergarteners, the subjects received four different tests: a syllable counting test (Liberman et al., 1974), a test of memory for phonetically confusable and phonetically nonconfusable word strings (Mann et al., 1980), the Corsi block test (Corsi, 1972), and the Peabody Picture Vocabulary Test (Dunn, 1959). As first graders, they again received the word-string test and the Corsi block test and were further given the Word Recognition and Word Attack subtests of the Woodcock Reading Mastery Test (1973). Materials for the experimental tests are described below.

Syllable counting test. Training and test materials for this test are described in full in Liberman et al. (1974) and are listed in Appendix A. The training materials consisted of four three-word items in which the first word has one syllable, the second has two syllables, and the third has three syllables (e.g., "but," "butter," "butterfly"). The test materials consisted of a randomized list of 42 common words, with one-, two-, and three-syllable words equally represented in random order.

Word-string memory test. Materials for this test consisted of 16 different word strings, each of which contained four words. Eight of the strings contained words that rhymed with each other (the phonetically confusable strings) and eight contained words that did not rhyme (the phonetically nonconfusable strings). Each of the eight phonetically confusable strings consisted of four one-syllable words drawn from the Thorndike and Lorge A and AA frequency class (Thorndike & Lorge, 1944). The four words rhymed with each other but were not semantically related. To construct the phonetically nonconfusable strings, the phonetically confusable strings were divided into two sets of four strings each, and the words within each set were then randomized so as to form four phonetically nonconfusable strings in which none of the four words rhymed. From the total corpus of phonetically confusable and phonetically nonconfusable word strings, we then composed two lists (Lists A and B) of eight word strings each. These lists are given in Appendix B. Each list contained one of the two sets of phonetically confusable strings interspersed with the complementary set of phonetically nonconfusable word strings. Thus, those words that occurred as part of a rhyming string in one list occurred as part of a nonrhyming string in the other list, and no word occurred twice within a single list.

Corsi block test. Materials for this test, as described in Milner (1972), consist of a set of nine 3 cm wooden cubes, mounted onto a 28 by 23 by 1 cm base. The cubes are placed in a semi-random array and the entire apparatus is painted black so as to eliminate all surface detail. Identifying numbers, which are painted on one side of the base, are visible only to the examiner.

Procedure

For the kindergarten phase of testing, two 20-minute sessions were required, whereas first-grade testing was accomplished in a single 30-minute session. All children were tested individually and received the tests in the same order. Standard procedures were followed for administering the Peabody and the Woodcock tests; procedures for the other tests are given below.

Syllable counting test. The procedure for this test has been described in Liberman et al. (1974). Under the guise of a "tapping game," the child was required to repeat a word spoken by the examiner and to indicate the number (from one to three) of syllables in that word by tapping a small wooden dowel on the table. During training, each of the training sets of three words was first demonstrated by the experimenter in order of increasing syllables. When the child was able to repeat and correctly tap each item in the set in the order demonstrated during initial presentation, the items of the triad were then presented in scrambled order without prior demonstration. The child's tapping was corrected as needed. In the test trials that followed, each word was given without prior demonstration and corrected by the experimenter as needed. Testing continued through all 42 items. Two scores were computed for each child: a pass/fail score based on whether or not a child had at any point during testing performed six consecutive items correctly, and an error score reflecting the total number of words missed.

Word-string memory test. The examiner began this test by telling the child that some words would be spoken, one at a time, and that the child's job was to listen carefully and try to repeat the entire word string in the order heard. A practice item consisting of the string "cat, house, foot, tree" was then given, the words being spoken at the rate of one per second. A second practice item followed, consisting of the sequence "egg, brush, leaf, dog." At this point, actual testing began. The child now listened to a loudspeaker that played a taped sequence of the examiner saying the test word strings. The delivery rate was one word per second. The tape was stopped after each word string to permit the child to respond, and all responses were immediately transcribed and also recorded for later re-analysis. During kindergarten testing, the subjects heard the two lists in different sessions; as first graders, they completed both lists in a single session, separated by a 20-minute break.

In scoring the children's responses, phonetically confusable and phonetically nonconfusable strings were treated separately. For each string, an error score was computed by counting a word as incorrectly recalled if it was omitted or if it occurred in the improper sequence relative to the first correctly-recalled word that preceded it. Only the first four responses given to each string were considered. Since there were eight strings in the phonetically confusable and phonetically nonconfusable sets, the total possible error score was 32 for each set. Whereas scores on individual strings were entered into analyses of covariance, total error scores were entered into the multiple regression.

Corsi block test. Seated opposite the child and facing the numbered side of the base, the experimenter explained that some blocks would be tapped, one at a time. The child was instructed to watch the examiner tap the blocks and then to try to touch the same blocks in the same order. The experimenter used a randomized digit sequence as a guide to which block sequences to touch, and tapped each block at the rate of one per second. As the subject responded, the sequence was recorded in terms of the corresponding digits. Eight practice items were given first, which consisted of four two-block sequences and four three-block sequences. The test followed and consisted of eight items: four four-block sequences and four five-block sequences. Response feedback was not provided during testing. In scoring each child's responses, an error score was computed for each test sequence. A block was considered incorrectly recalled if it was omitted or recalled in the improper sequence relative to the first correct block that preceded it. Error scores were then summed for the eight test sequences, with the maximum score being 36.

RESULTS

In assessing the results of our study, the first question of interest was whether performance on any of our tests would be significantly related to reading ability in the first grade. We began answering this question by dividing the children into three reading groups according to their first-grade teachers' recommendations. There were 26 good readers, 19 average readers, and 17 poor readers. As a means of corroborating these ratings, we next computed the sum of each child's score on the Word Attack and Word Recognition subtests of the Woodcock. We found the mean sum of scores for good readers (109.1) to be significantly higher than that of average readers (65.1), $t(43) = 8.85$, $p < .005$, which was in turn significantly higher than that of the poor readers (34.5), $t(34) = 6.75$, $p < .005$. Children in the three different reading groups did not, however, differ in age or in IQ.

Having thus subdivided our subjects according to reading ability, we conducted a series of analyses of covariance which adjusted for any effects of age and IQ. We examined whether reading level was significantly related to performance on any of our three tests--the syllable counting test, the word-string memory test, and the Corsi block test.

Syllable counting. With regard to the syllable counting test, of the 26 children classified as good readers in the first grade, 85% had reached the criterion of six consecutive items correct as kindergarteners. In contrast, only 56% of the average readers and only 17% of the poor readers had done so. An analysis of covariance performed on children's error scores confirms the significance of these differences, $F(2,56) = 7.98$, $p < .001$.

Word-string memory. Children's mean error scores on the word-string memory test are given in Table 1, with scores obtained during the kindergarten

Table 1

Mean Error Scores of Good, Average and Poor Readers on Memory Tasks: A Longitudinal Study (IQ Determined in Kindergarten, Reading Achievement in First Grade).

Reading Ability	Grade Level	Word-string Memory Max=32 Nonrhyming Word Strings	Word-string Memory Max=32 Rhyming Word Strings	Corsi Block Memory Max=32
Good Readers				
N=26	KDGN	8.1	13.4	8.4
IQ 114.7	1st Grade	5.5	12.1	8.7
Average Readers				
N=19	KDGN	12.8	15.4	9.0
IQ 114.7	1st Grade	9.2	11.3	8.1
Poor Readers				
N=17	KDGN	13.2	15.0	10.1
IQ 115.5	1st Grade	13.7	12.7	10.1

phase of testing separated from those obtained in the first-grade phase. In general, children made more errors as kindergarteners, $F(1,58) = 30.28$, $p < .001$. On the average, they also made more errors on the phonetically confusable word strings than on the nonconfusable ones, $F(1,58) = 76.9$, $p < .001$.

Differences among the three reading groups are most important to our predictions. On the average, the number of errors was inversely related to a child's reading ability, $F(2,56) = 6.29$, $p < .004$. In addition, as we had discovered in the past, the extent of difference among children in the three reading groups was greater in the case of phonetically nonconfusable word strings than in the case of confusable ones, $F(2,58) = 14.0$, $p < .001$. This interaction reflects the fact that good readers were more penalized by the presence of phonetic confusability than were children in the other two reading groups.

It is clear from Table 1 that as first graders, good readers made significantly fewer errors than poor readers. This would be expected, of course. One is also not surprised to find, in addition, that differences in the verbal memory performance of the three reading groups were greater when the children were first graders than when they were kindergarteners, $F(2,58) = 4.5$, $p < .02$. However, it is particularly important, in our view, to note that the differences were nonetheless present before children entered the first grade. As kindergarteners, the future good readers had made significantly fewer errors, in general, than poor readers, $t(41) = 4.52$, $p < .001$; as first-graders, these differences remained, $t(41) = 2.56$, $p < .02$. Average readers fell somewhere in between--closer to poor readers in kindergarten and closer to good readers in first grade.

As to phonetic confusability, when they were kindergarteners, both the future good and average readers had made significantly more errors on confusable strings than on nonconfusable ones [$t(25) = 5.8$, $p < .001$ for the good readers; $t(18) = 2.7$, $p < .05$ for the average ones], whereas poor readers showed the same level of performance on both string types [$t(16) = 1.42$, $p > .10$]. As first graders, the good and average readers again made more errors on phonetically confusable strings [$t(25) = 9.6$, $p < .001$ and $t(18) = 2.23$, $p < .05$], whereas poor readers actually made an equivalent number of errors on the two word-string types [$t(16) = 1.01$, $p > .10$].

Corsi blocks. Mean scores on the Corsi block test are also displayed in Table 1. As can be seen in that table, any differences among children in the three reading groups were minimal. Analysis of covariance reveals no significant effect of reading level, or of age at testing. Although poor readers averaged slightly lower than other children, a series of t-tests revealed that the scores of poor readers are equivalent to those of children in the other two reading groups.

Regression analysis. As a final and alternative means of analyzing the data, we computed linear regressions of reading ability (as measured by the sum of Woodcock scores) onto the scores of our various experimental tests. Two separate regressions were computed, one for results obtained during kindergarten testing, and one for those obtained during first grade testing.

In the case of kindergarten testing, two scores were significantly correlated with reading ability at the .01 level--syllable counting [r(58) = .40], and memory for the phonetically nonconfusable word strings [r(58) = .39]. Performance on the phonetically confusable words was correlated with reading ability at the .05 level [r(58) = .33]. We were also interested to discover that performance on syllable counting was somewhat correlated with memory for the phonetically nonconfusable word strings, r(58) = .26, $p < .05$. (As might be expected, performance on the nonconfusable word strings was also correlated with that on the confusable ones, r(58) = .66, $p < .001$.) Taken together, error scores on syllable counting and memory for phonetically nonconfusable word strings account for 24% of the variance in reading scores; each uniquely accounts for 9% of the variance. The analagous regression computed on the first-grade scores upheld the kindergarten results, revealing a strong correlation between reading ability and performance in memory for the phonetically nonconfusable word strings, r(58) = .61, $p < .001$. (Once again, performance on the nonconfusable strings also correlated with that on the confusable ones, r(58) = .52, $p < .01$.) Performance on the phonetically nonconfusable word strings accounted for 40% of the variance in reading ability, 25% of which was unique.

Sex differences. Although our experimental population contained an equal number of boys and girls, the two sexes were not equally distributed among our three reading groups. Of the good readers, 64% were girls, whereas only 35% of the poor readers were girls. Yet, within each reading group, the performance of boys and girls in that group was similar. Although more girls were good readers, their performance was not qualitatively different from boys who were good readers; similarly, although more boys were poor readers, their performance was not qualitatively different from girl poor readers. For a further discussion of sex differences in these data, see Liberman and Mann (1981).

DISCUSSION

Our hypotheses about the interrelationships among beginning reading ability, phonological awareness, and verbal short-term memory were initially motivated by theoretical considerations about the relation of language skill and reading. They were substantiated in experiments that examined either the association between reading ability and phonological awareness, or between reading ability and verbal short-term memory in first- or second-grade children. Now, the results of our longitudinal study show that phonological awareness and verbal short-term memory do more than correlate with early reading ability. They reveal that, among kindergarteners, the adequacy of these two language skills may presage future reading ability in the first grade. They also suggest at least a moderate correlation between phonological awareness and verbal short-term memory.

Some of our earliest work had revealed that phonological awareness is associated with reading success (Liberman, 1973; Liberman et al., 1974). Phonological awareness, as measured by a child's ability to count phonemes in a spoken utterance, was found to predict reading success in the first grade. That is, children who failed a phoneme counting test, analagous to the present syllable counting test, were highly likely to become the poorer readers of their classrooms. The results of the present study reveal that the ability to count syllables in spoken utterances can also be a predictor of reading success. Moreover, syllabic awareness has the advantage of being less easily confounded by reading instruction. This latter fact can be seen in a recent Belgian study that compared the phonological awareness of children receiving a "phonics" type of reading instruction with that of children receiving a "whole-word" type of instruction (Alegria et al., in press). The "phonics" group showed a greater awareness of phonemic structure than did the "whole-word" group (60 percent correct as opposed to a mere 16 percent correct). The two groups were not very different, however, in their awareness of syllable structure (72 percent correct as opposed to 63 percent correct). Thus, differential reading instruction at the first-grade level apparently has a marked effect on phonemic awareness but not on syllabic awareness.

So much for phonological awareness. In our previous work, as we have noted earlier, we had also found verbal short-term memory skill to be related to beginning reading ability. As compared to poor beginning readers, the good readers were more able to remember a string of letters (Liberman et al., 1977; Shankweiler et al., 1979), a string of words (Mann et al., 1980), and even the words of a sentence (Mann et al., 1980), perhaps because they make more effective use of phonetic representation in short-term memory. The present study confirms this association in the case of first-grade children, but further reveals that the advantage in verbal short-term memory skill actually preceded first-grade reading success. Among the children we tested, kindergarteners who did well in repeating the word strings were likely to become the better readers of their first-grade classrooms. In addition, the future good readers were showing evidence of relying on phonetic representation, as seen in their particular difficulty with repeating strings of phonetically confusable words. The future poor readers, on the other hand, were relatively tolerant of our manipulations of phonetic confusability, and the future average readers fell somewhere in between.

We should note that it was only the two language skills in our study that proved to relate to success in beginning reading. IQ scores in the range encountered in the normal classroom were not adequate predictors of reading success. Similarly, performance on the nonverbal short-term memory test also failed to differentiate poor beginning readers from the more successful readers in their classrooms. In the light of these findings, it would seem that our poor readers were not reading disabled because of a general intellectual deficiency, nor because they suffered from some general short-term memory deficiency, as has been suggested by some (Morrison, Giordani, & Nagy, 1977). Their problems appear, instead, to be related to language processing.

Suggestions for Kindergarten Screening

A primary contribution of this study, in our view, is to suggest that kindergarten-level performance on language-based tasks--a test of phonological awareness and a test of verbal short-term memory--may presage first-grade reading ability and might therefore be used as part of a kindergarten screening battery. It is true that performance on these tests accounts for only a quarter of the total variance in our subjects' reading ability. These tests would, therefore, not be capable of predicting differences within a group of good readers, for example. Nonetheless, the tests would be very useful in predicting the extremes of reading success in the first grade. That is, a kindergartener who does well on both syllable counting and verbal short-term memory has a significant likelihood of later becoming a successful beginning reader; a child who does poorly on both has a significant likelihood of later becoming a poor reader. That information is surely worth knowing as soon as possible, and anyone interested in screening children to find those at risk for reading problems might therefore do well to consider using these two easily administered tasks as part of a screening battery. The children who fell in the lower quartile of the class on one of these tasks, and certainly those who did so on both, might then be considered at risk.

The Corsi block test might be added as well, as a control for possible problems in attention span. Whereas a child who does poorly on the Corsi block test alone is not necessarily a candidate for possible reading problems, a child who does poorly on the Corsi block test _and_ on syllable segmentation _and_ on verbal short-term memory tests may have a language problem, but might also have an attentional deficit that could in itself be expected to lead to learning problems.

Although these tests may be sufficient for most screening purposes, other language-based tests might be considered as well. One that might be suggested is a test of rapid letter-naming ability. This would add a measure of speed of word retrieval to the other measures of language processing. Rapid automatized naming (RAN) of letters (Denckla & Rudel, 1976) has been found on numerous occasions to be related to reading ability. Blachman (1980) recently found that a test that included phoneme segmentation, a measure of verbal short-term memory, and RAN letter naming accounted for a large part of the variance in first-grade reading.

Implications for Prevention of Reading Problems

Having administered these tests to the kindergarteners and having thus identified those children at risk for reading problems, a teacher could then begin to direct efforts toward preventing future reading problems. As every teacher knows, it is one thing to screen for problems, but quite another to do something about them. A critical question, then, is what these tasks might tell us about the form that preventive efforts should take. They certainly suggest that the efforts should be language-based. Beyond that, what else can be said?

In earlier papers (Liberman & Shankweiler, 1979; Liberman, Shankweiler, Blachman, Camp, & Werfelman, 1980) some suggestions relating to the improvement of phonological awareness were outlined. We discussed several pre-reading techniques there that have been found to facilitate the awareness of the structure of spoken words that is so important for the development of proficiency in reading an alphabetic orthography. To begin with, teachers can use many indirect methods that manipulate phonological structure. For example, they can capitalize on some common forms of word play, such as teaching the children nursery rhymes, encouraging rhyming games that include nonsense words, and promoting "secret" languages such as "Pig Latin" and "Ubby-Dubby." Later, direct awareness training can be initiated. Since the word and the syllable are more readily extracted from the speech stream than the phoneme, direct phonological training would best proceed from word awareness to syllable awareness and finally to phoneme awareness. To make the word explicit, we favor counting games such as those suggested by Engelmann (1969) in which the teacher instructs the child to repeat and then to count the words in sentences, beginning with such simple statements as "John is happy," to which complexities are added as needed. To impart an awareness of syllabic structure, the elision task described by Rosner and Simon (1971) could then be employed. Children would, for example, be asked to "say 'cowboy' without the 'cow'." They could even be given explicit training in our own syllable-counting task. Finally, phonemic awareness could be introduced with the procedure of the Soviet psychologist Elkonin (1973).

In Elkonin's procedure, the child is presented with a line drawing of an object that he or she knows well. Below the picture is a rectangle divided into sections corresponding to the number of phonemes in the pictured word. The child is taught to say the word slowly, putting a counter in the appropriate section of the diagram as he or she pronounces the word. After playing this "game" with many different pictured words until the diagram is no longer necessary, the child is introduced to the concept of vowels and consonants. At this time, one color of counter is used for vowels and another for consonants. Finally, proceeding with a single vowel at a time, graphemes are added to the counters. The child then masters the names and sounds of the five short-vowel letters, after which consonant graphemes are gradually introduced. There are many pedagogical virtues to this procedure. First, the diagram provides a linear visuospatial structure to which the auditory-temporal sequence of the word can be related, thus reinforcing the key idea of successive segmentation of the phonemic components of words--an idea intrinsic to an alphabetic system, and one best learned as soon as possible. Second, the actual number of segments is provided for the child, so that uninformed guessing of the number of components is not necessary. Finally, the picture keeps the word in front of the child during analysis so that there is minimal stress on verbal short-term memory--something that we already know will be a problem for many children.

That brings us to the question of how to improve verbal short-term memory skill--or whether it can be improved. It could well be that the problems some children have with verbal short-term memory are the consequences of a maturational lag (Satz, Taylor, Friel, & Fletcher, 1978). If so, then we might expect to see some gradual improvement as the children progress through school. It has been reported (Holmes & McKeever, 1979; McKeever & Van Deventer, 1975), however, that a verbal memory deficit characterizes adolescent poor readers, just as it characterized the poor beginning readers we have tested. Perhaps future longitudinal studies will shed more light on this issue.

For the moment, we do not know whether or not poor readers will outgrow their language problems. In fact, it is at least possible that their deficits are of a more permanent nature. In that case, the deficiencies we observe among some poor beginning readers could be symptoms of a "subclinical" aphasia that is due to a subtle deficit in the left or language-dominant hemisphere. There are, after all, some interesting parallels between poor beginning readers and adults who have suffered damage to their language-dominant hemisphere. Verbal short-term memory, for example, is often deficient among adult aphasics, whereas Corsi block performance is not (Corsi, 1972; Milner, 1972). Further clarification of the similarities and dissimilarities between early reading disability and acquired aphasia is a project that concerns us at present.

As for remediation of verbal short-term memory problems, we do not have as clear an idea of how to answer this question as we did for phonological awareness. If the problem is not simply ameliorated with time, then we can only suggest practice, practice, and more practice. Having children repeat spoken sentences may be a good idea--and that is something that the Engelmann procedure will require anyway. Learning to repeat nursery rhymes and other poetry may help, and certainly will not hurt. Increased emphasis on language arts in general, and on grammatical skills in particular, may well serve to enhance verbal memory by providing an emphasis on the structural aspects of language. In our view, it is not beyond the realm of possibility that the present epidemic of illiteracy reflects to some degree the decreased emphasis on memorization, recitation, sentence parsing, and rhetoric. Here again, further research may provide some answers.

REFERENCES

Alegria, J., Pignot, E., & Morais, J. Phonetic analysis of speech and memory codes in beginning readers. *Memory & Cognition*, in press.

Baddeley, A. D. The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review*, 1978, *85*, 139-152.

Blachman, B. *The role of selected reading related measures in the prediction of reading achievement*. Unpublished doctoral dissertation, University of Connecticut, 1980.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, *55*, 75-84.

Corsi, P. M. *Human memory and the medial temporal region of the brain*. Unpublished doctoral dissertation, McGill University, 1972.

Denckla, M. B., & Rudel, R. G. Rapid automatized naming (R.A.N.): Dyslexia differentiated from other learning disorders. *Neuropsychologia*, 1976, *14*, 471-479.

Dunn, L. L. *Peabody picture vocabulary test*. Circle Pines, Minn.: American Guidance Service, 1959.

Elkonin, D. B. USSR. In J. Downing (Ed.), *Comparative reading*. New York: Macmillan, 1973.

Engelmann, S. *Preventing failure in the primary grades*. Chicago, Ill.: Science Research Associates, Inc., 1969.

Fox, B., & Routh, D. K. Analyzing spoken language into words, syllables, and phonemes: A developmental study. *Journal of Psycholinguistc Research*, 1975, *4*, 331-342.

Goldstein, D. M. Cognitive-linguistic functioning and learning to read in preschoolers. *Journal of Educational Psychology*, 1976, *68*, 680-688.

Golinkoff, R. Phonemic awareness skills and reading achievement. In F. B. Murray & J. J. Pikulski (Eds.), *The acquisition of reading: Cognitive, linguistic and perceptual prerequisites*. Baltimore, Md.: University Park Press, 1978.

Holmes, D. R., & McKeever, W. F. Material specific serial memory deficit in adolescent dyslexics. *Cortex*, 1979, *15*, 51-62.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. Memory for item order and phonetic recoding in the beginning reader. *Journal of Experimental Child Psychology*, 1981, *32*, 474-484.

Liberman, A. M., Mattingly, I. G., & Turvey, M. Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), *Coding processes and human memory*. Washington, D.C.: Winston and Sons, 1972.

Liberman, I. Y. Basic research in speech and lateralization of language: Some implications for reading disability. *Bulletin of the Orton Society*, 1971, *21*, 7-87.

Liberman, I. Y. Segmentation of the spoken word. *Bulletin of the Orton Society*, 1973, *23*, 65-77.

Liberman, I. Y., Liberman, A. M., Mattingly, I. G., & Shankweiler, D. Orthography and the beginning reader. In J. F. Kavanagh & R. Venezky (Eds), *Orthography, reading, and dyslexia*. Baltimore, Md.: University Park Press, 1980.

Liberman, I. Y., & Mann, V. A. Should reading instruction and remediation vary with the sex of the child? In N. Geschwind, M. Albert, A. Galaburda, & N. Gartrell (Eds.), *The significance of sex differences in dyslexia*. Towson, Md.: Orton Society, 1981.

Liberman, I. Y., Mann, V. A., Shankweiler, D., & Werfelman, M. Children's memory for recurring linguistic and nonlinguistic material in relation to reading ability. *Cortex*, in press.

Liberman, I. Y., & Shankweiler, D. Speech, the alphabet, and teaching to read. In L. Resnick & P. Weaver (Eds.), *Theory and practice of early reading*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1979.

Liberman, I. Y., Shankweiler, D., Blachman, B., Camp, L., & Werfelman, M. Steps toward literacy. Report prepared for Working Group on Learning Failure and Unused Learning Potential, President's Commission on Mental Health, Nov. 1, 1977. In P. Levinson & C. H. Sloan (Eds.), *Auditory processing and language: Clinical and research perspectives*. New York: Grune & Stratton, 1980.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology*, 1974, *18*, 201-212.

Liberman, I. Y., Shankweiler, D., Liberman, A. M., Fowler, C., & Fischer, F. W. Phonetic segmentation and recoding in the beginning reader. In A. S. Reber & D. L. Scarborough (Eds.), *Toward a psychology of reading: The proceedings of the CUNY Conference*. Hillsdale, N.J.: Erlbaum, 1977.

McKeever, W. F., & Van Deventer, A. D. Dyslexic adolescents: Evidence of impaired visual and auditory language processing associated with normal lateralization and visual responsivity. *Cortex*, 1975, *11*, 361-378.

Mann, V. A., Liberman, I. Y., & Shankweiler, D. Children's memory for sentences and word strings in relation to reading ability. *Memory & Cognition*, 1980, *8*, 329-335.

Morais, J., Carey, L., Alegria, J., & Bertelson, P. Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 1979, *7*, 323-331.

Milner, B. Disorders of learning and memory after temporal-lobe lesions in man. *Clinical Neurosurgery*, 1972, *19*, 421-446.

Morrison, F. J., Giordani, B., & Nagy, I. Reading disability: An information processing analysis. *Science*, 1977, *196*, 77-79.

Rosner, J., & Simon, D. P. *The auditory analysis test: An initial report*. *Journal of Learning Disabilities*, 1971, *4*, 40-48.

Satz, P., Taylor, H. G., Friel, J., & Fletcher, J. Some developmental and predictive precursors of reading disabilities: A six-year follow-up. In A. L. Benton & D. Pearl (Eds.), *Dyslexia: An appraisal of current knowledge*. New York: Oxford University Press, 1978.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. The speech code and learning to read. *Journal of Experimental Psychology: Human Learning and Memory*, 1979, *5*, 531-545.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, 1944.

Wiig, E. H., & Semel, E. M. *Language disability in children and adolescents*. Columbus, Ohio: Charles E. Merrill, 1976.

Woodcock, R. W. *Woodcock Reading Mastery Tests*. Circle Pines, Mn.: American Guidance Services, Inc., 1973.

APPENDIX A

Materials for Syllable Counting Test.

Training trials

1. but
   butter
   butterfly

2. tell
   telling
   telephone

3. doll
   dolly
   lollipop

4. top
   water
   elephant

Test List

1. popsicle
2. dinner
3. penny
4. house
5. valentine
6. open
7. box
8. cook
9. birthday
10. president
11. bicycle
12. typewriter
13. green
14. gasoline
15. children
16. letter
17. jump
18. morning
19. dog
20. monkey
21. anything
22. wind
23. nobody
24. wagon
25. cucumber
26. apple
27. funny
28. boat
29. father
30. holiday
31. yellow
32. cake
33. fix
34. break
35. overshoe
36. pocketbook
37. shoe
38. pencil
39. superman
40. rude
41. grass
42. fingernail

APPENDIX B

Materials for Word-string Memory Test

List A

1.	(nonrhyming)	bee	hair	gate	head
2.	(nonrhyming)	chair	plate	knee	bed
3.	(rhyming)	nail	tail	sail	mail
4.	(rhyming)	fly	tie	pie	sky
5.	(nonrhyming)	red	tree	bear	state
6.	(rhyming)	meat	heat	feet	street
7.	(nonrhyming)	thread	pear	weight	key
8.	(rhyming)	brain	train	chain	rain

List B

1.	(rhyming)	pear	bear	chair	hair
2.	(nonrhyming)	tie	rain	heat	tail
3.	(rhyming)	state	plate	gate	weight
4.	(nonrhyming)	train	sky	feet	sail
5.	(rhyming)	bee	tree	knee	key
6.	(nonrhyming)	meat	nail	fly	brain
7.	(rhyming)	bed	head	thread	red
8.	(nonrhyming)	mail	chain	pie	street

INITIATION VERSUS EXECUTION TIME DURING MANUAL AND ORAL COUNTING BY STUTTERERS*

Gloria J. Borden+

Abstract. Severe stutterers were found to be significantly slower than control subjects in performing a speech counting task that was judged to be fluent and in silently counting on their fingers. For both tasks the time taken to execute the series accounted for more of the difference between severe stutterers and controls than the time taken to prepare and initiate the series. Mild stutterers were not significantly slower than controls on either task.

The main purpose of the experiment, from which this paper is the first report, was to examine the interactions of respiratory, laryngeal, and supralaryngeal movements of stutterers and their controls during speech. A second purpose was to examine finger movements in a nonspeech serially-ordered task in order to find out if differences between stutterers and controls extend beyond the speech mechanisms. The final purpose was to study the interactions between the manual and oral movements when engaged in a common task. To make these comparisons, the task of counting was chosen, since it is a serially-ordered event, and subjects can count aloud, silently count on their fingers, and simultaneously count aloud and manually.

The present paper is a report on the timing of intervals measured during the speech-alone and fingers-alone conditions. A reaction time paradigm was used to maximize the probability that stuttering would occur in the laboratory setting and to examine the role of planning in the execution of the tasks. Of special interest was the comparison of the timing of intervals for the perceptually fluent utterances of stutterers with the utterances of the normal speakers.

Recent investigations into the timing of motor responses of stutterers have indicated that, as a group, they may be motorically slower than nonstutterers even during their seemingly fluent utterances. Slower speech movements have been measured from x-ray films of articulators (Zimmermann, 1980a), inferred either from slower formant changes (Starkweather & Myers,

---

*This paper is under consideration for publication in the Journal of Speech and Hearing Research.

+Also Temple University, Philadelphia, Pennsylvania.

Acknowledgment. Stutterers were referred by Bernard Stoll and Arlyne Russo, who also tested them for severity. The glove for recording digital contacts was constructed by David Zeichner, and the program for displaying the visual signals was written by Edward Wiley with the assistance of Donald Hailey. Technical assistance was provided by Richard Sharkany. The Electroglottograph was borrowed from Temple University and the experiment conducted at Haskins Laboratories. Cynthia Keely provided a reliability check of the interval measurements. Helpful comments were offered by Katherine S. Harris and J. A. Scott Kelso. This study was funded in part by NIH grant NS 13617.

1979) or increased phonatory reaction time (Adams & Hayden, 1976; Starkweather, Hirschman, & Tannenbaum, 1976), or observed in increased latency of muscle activity (McFarlane & Prins, 1978). It has further been suggested that stutterers may be slower than normal to perform manual as well as speech motor acts (Luper & Cross, Note 1). Other studies have failed to find evidence of a significant difference in manual latency between stutterers and controls, but found stutterers to be slower in producing the sounds of speech (Prosek, Montgomery, Walden, & Schwartz, 1979; Reich, Till, & Goldsmith, 1981).

Most of the investigations comparing the latency of stutterers and controls have focused on the time between a signal to respond and the onset of the response. This interval may be considered the *initiation* time, an interval that includes pre-motor planning and motor initiation. It seemed interesting to include in such studies the interval that may be termed *execution* time--the interval between the first and last event in a serially-ordered response. Since stuttering episodes predominate at the onset of words and phrases (Bloodstein, 1975), initiation seems to present a greater problem for stutterers than continuing execution. Both initiation and execution measures were therefore included in the design to permit comparison of the two intervals.

Further, it is possible to evaluate the importance of pre-movement preparation by comparing a condition in which the response is known ahead of the signal to respond (*delayed response condition*) with a condition in which the expected response is displayed simultaneously with the signal to respond (*immediate response condition*) (Ostry, 1980). If the response is brief and the expected response is known one second before the signal to respond, certain preparatory events may be presumed to have occurred before the signal to respond, such as perceiving the response to be executed and priming several groups of muscles for the coming activity.

The investigations of manual response time in stutterers cited above used a key-press response. Such a response requires a simple ballistic movement that is not completely analogous to the coordination of different muscle groups necessary for speech. Counting on one's fingers requires that many groups of muscles work together. Further, pressing an external object such as a button or a keyboard seems less like speech than does counting on one's own fingers, a situation in which the "targets" are intrinsic to the counter. The rationale for choosing finger counting was based on the fact that it is a serially-ordered response, self-contained, and requires complex motor coordination.

Thus, the present study compares the initiation time versus execution time measured from the responses of stutterers and their controls in two serially-ordered tasks: counting four-digit numbers aloud and on fingers. It was also designed to evaluate the role of planning by including an immediate-response condition and a delayed-response condition. The primary purpose of this part of the experiment was to compare the initiation and execution intervals in the seemingly fluent utterances of stutterers with the same intervals in the utterances of the controls. A secondary purpose was to compare stutterers with controls in the times taken to initiate and execute the finger counting task. Of overall interest was whether stutterers are generally slower than normal in the performance of motor tasks.

METHOD

Subjects

Eight adult stutterers (7 male and 1 female) ranging in age from 21 to 48 were matched in pairs by sex, age, and general educational/occupational level with eight normal speakers ranging in age from 20 to 45. Mean age for the experimental group was 33 and for the control group was 32. College students, teachers, blue collar workers, and professionals were represented in both groups. Subjects were bimodally distributed in terms of the severity of their stuttering. Four of the stutterers were rated as mild and four as severe, according to the Stuttering Severity Index (Riley, 1972), the reading and conversational parts of the Stuttering Interview (Ryan, 1974), and subjective judgments of two speech pathologists (see Table 1).

The stutterers were recruited through the assistance of speech pathologists in the New Haven area: their controls were volunteers who matched the experimental subjects in age (within 5 years), sex, and general background. All subjects reported themselves to be right-handed in writing, throwing, hammering, and cutting with scissors.

Tasks

The two response tasks reported here are speech counting and finger counting. The speech counting task involved reading aloud a digital display of ten different sequences of the digits 2, 3, 4, and 5. Each sequence appeared twice: once simultaneously with a tonal response signal (*immediate* condition) and once 1 sec before the sounding of the signal to respond (*delayed* condition). The 20 items were randomized. Any visual or auditory evidence of stuttering was marked during the experiment, as well as any errors in counting. Any visual sign of struggle or effort in facial or body movements was noted, as was any auditory sign of hesitation, repetition, or prolongation. Although this paper emphasizes an analysis of the perceptually fluent utterances of the stutterers, another purpose of the experiment was to compare fluent with stuttered utterances. Thus, for all subjects the speech counting task was followed by the finger counting task in order to maximize the probability that stuttering would occur. This poses a problem for any comparison of speech with nonspeech conditions, since order may have an effect. It was a risk felt to be worth taking, however, and one can still compare stutterers with controls on the manual task. A different randomization of the same 20 items was presented for the manual task, and the subjects silently counted on their fingers by contacting index finger and thumb for the number 2, middle finger and thumb for 3, ring finger and thumb for 4, and little finger and thumb for 5. All finger counting was done on the right hand. Instructions were read to each subject before a practice set of 12 sequences. Instructions included a warning to wait for the tone before responding and to count as quickly as possible without sacrificing accuracy. Practice was given on both tasks. None of the practice sequences appeared on the tests.

TABLE 1

Subject identification, sex, age, and judged severity of stuttering.

Experimental Group

1.	JP	M	48	severe
2.	DE	M	22	severe
3.	DA	M	31	severe
4.	LB	M	44	mild
5.	DL	F	30	severe
6.	MA	M	26	mild
7.	GV	M	41	mild
8.	SL	M	21	mild

$\bar{X} = 33$

Control Group

1.	FS	M	45
2.	TS	M	22
3.	SB	M	30
4.	EG	M	43
5.	NM	F	32
6.	JL	M	29
7.	AL	M	36
8.	DR	M	20

$\bar{X} = 32$

Instrumentation

The program presenting the test sequences was run on a microcomputer (Integrated Computer Systems). For each sequence, a visual warning signal was followed by a variable interval (300, 400, or 500 msec), after which the 4-digit display appeared. The tone signalling the subject to respond was either simultaneous with the display or delayed 1 sec after the display. Presentation of the next test sequence was experimenter-controlled to allow for subject differences in response time.

An electroglottograph (F-J Electronics ApS) recorded rapid changes in impedance by high pass filtering (25 Hz-10 kHz) the overall changes in impedance of an imperceptible signal transmitted across the larynx at the level of the vocal folds. The onset of these rapid oscillations was abrupt and unambiguous and served to signal the onset of voicing during the speech task.

A special glove made of thin cotton was constructed for the right hand with circles of thin (.0015 inch) brass attached to each finger pad and a larger thimble-shaped contact surface attached to the thumb. Each contact produced a different voltage. These signals served to represent the onset of each digital contact during finger counting.

The electroglottograph and glove signals, along with the speech acoustic signal, were recorded on an EMI SE-7000 FM tape recorder. Other movement indices not included in the present analysis include chest and abdominal wall movement and lower lip movement. The interaction of these movements will be described in a future report.

Measurement of Intervals

Visicorder recordings of the physiological and acoustic signals recorded on FM tape were produced for each subject. Onset of voicing as inferred from the laryngographic signal and onset of finger contacts were marked by the experimenter. All subject errors were omitted from the measured data, including counting confusions and responses started before the signal to respond. These errors were categorized, however, for analysis of any speed-accuracy tradeoff. Dysfluencies were classified separately from fluent utterances for measurement. Dysfluencies included those evident in the movement traces as well as any auditory or visual indications of stuttering identified during the tests. For example, the appearance of rapid fluctuations in laryngeal impedance during the silence before speech was classified as dysfluent. Thus, in an utterance classified as "fluent," the subject gave no visual sign of struggle in facial or body movements, the speech had to be free from any auditory sign of hesitations, repetitions, or prolongations, and the physiological traces examined later had to be free from abnormal perturbations or oscillations. Measures were made in milliseconds from the response signal to the onset of the first response (initiation time) and from the onset of the first response to the onset of the last response (execution time). Measurements made by the experimenter were repeated by a research assistant and any discrepancy over 10 msec was remeasured by both for consensus.

Analysis of the Data

For each subject, means and standard deviations were computed for initiation time in the delayed condition, initiation time in the immediate condition, execution time in the delayed condition, and execution time in the immediate condition. For the speech task, means were computed separately for the utterances of the control subjects, the perceptually fluent utterances of the stutterers, and the dysfluent utterances of the stutterers. Stuttered utterances differed sufficiently from the fluent and control utterances that the need for a test of significance was precluded. The $\underline{t}$ test was used to test the significance of differences in interval times between the fluent tokens of stutterers and those of nonstutterers, and between finger counting by stutterers and their controls.

RESULTS

As noted above, the purpose of this portion of the study was to compare the interval times in the initiation and execution of the seemingly fluent utterances of stutterers with those of their controls during the speech counting task, and to compare the comparable interval times of the two groups in the finger counting task.

Speech Task

The fluent utterances of the stutterers were on the average about 20% slower than controls in the intervals measured for the speech task, while the stuttered tokens were about 178% slower, on average, than normal. Table 2 summarizes the means and standard deviations of initiation and execution times for each subject in both delayed and immediate response conditions. Averages are based on the measures from eight controls (C), the fluent tokens of six stutterers (F), and the dysfluent tokens of four stutterers (S). Two of the stutterers were dysfluent on all tokens, two were fluent for part and dysfluent for part, and four were judged fluent for the complete task. Fluent utterances were those in which the speaker sounded and looked fluent to the experimenter and there was no evidence of dysfluency (abnormal perturbations or tremor) on the physiological traces as observed on the Visicorder records. Table 2 shows that when subjects knew the series of numbers one second ahead (delayed condition), initiation time was reduced compared to the immediate-response condition. This advantage did not extend into the execution times for the remaining numbers in the series, however, for the control sample or for the fluent tokens of the stutterers. On the other hand, when averaged, the advantage of the delay did extend into the execution of the series in the dysfluent tokens of the stutterers.

There was a more extensive overlap of stutterers with controls in initiation time of fluent utterances than there was for execution time. The difference was significant on a $\underline{t}$ test for unequal n's between the fluent tokens of stutterers (n=6) and normals (n=8) in the time taken to execute the series ($\underline{t}(12) = 1.99$, $\underline{p} < .05$ delayed; $\underline{t}(12) = 2.23$, $\underline{p} < .025$ immediate), but there was not a significant difference in initiation time. The time difference is not due to a difference in strategy, which would have resulted in different numbers of errors in the two groups. An analysis of the errors

SPEECH COUNTING

$\bar{X}$ and (SD) in msec.

	Subjects	Delayed Initiation	Immediate Initiation	Delayed Execution	Immediate Execution
Dsyfluent Tokens	1. S	1911 (653)	2881 (624)	2094 (656)	2760 (1081)
	2. S	1294 (258)	1208 (311)	2677 (501)	2337 (436)
	*3. S	1245 (456)	1408 (267)	1150 (190)	2402 (998)
	+4. M	1070 (129)	1213 (152)	1126 (79)	1493 (399)
	Grand $\bar{X}$	1380 (318)	1678 (700)	1762 (657)	2248 (465)
Fluent Tokens	*3. S	530 (28)	804 (151)	1015 (35)	958 (36)
	+4. M	732 (94)	1033 (101)	1148 (153)	1064 (52)
	5. S	419 (48)	610 (69)	817 (49)	823 (87)
	6. M	403 (62)	552 (86)	776 (139)	812 (94)
	7. M	597 (98)	1110 (156)	714 (55)	719 (59)
	8. M	454 (95)	701 (84)	783 (120)	758 (90)
	Grand $\bar{X}$	523 (115)	802 (207)	876 (154)	856 (119)
Controls	1. C	361 (60)	587 (28)	696 (55)	700 (29)
	2. C	405 (51)	579 (142)	652 (42)	633 (45)
	3. C	470 (122)	673 (71)	608 (41)	624 (41)
	4. C	532 (114)	780 (102)	693 (64)	667 (71)
	5. C	486 (56)	586 (38)	759 (96)	712 (63)
	6. C	641 (110)	711 (92)	907 (77)	902 (75)
	7. C	469 (101)	708 (54)	634 (30)	638 (34)
	8. C	396 (53)	562 (43)	853 (70)	828 (42)
	Grand $\bar{X}$	470 (83)	648 (75)	725 (100)	713 (94)

C	Control
S	Severe
M	Mild
* +	Same subjects

Table 2. Means and standard deviations of speech intervals in milliseconds. Experimental subject 3 provided 6 fluent tokens and 14 dysfluent tokens, and experimental subject 4 provided 10 fluent tokens, 9 dysfluent tokens, and 1 discarded error.

excluded from the data revealed that only three of the control subjects and three of the stutterers made errors. The average number of errors among the three control subjects who made errors was 2 of the 20 utterances, while the average number of errors for the three stutterers who made errors was 1.7 out of 20. Most of the errors were early starts. Thus, accuracy was comparable in the two groups.

When the stutterers are grouped according to severity, four are rated as mild and four as severe. Two of the severe stutterers and all four of the mild stutterers produced fluent tokens on the speech counting task. Comparing each subject within each group with his or her individualized control in age, sex, and status, a different picture from that of the pooled data emerges (Figure 1). The left side of Figure 1 illustrates the extent of the overlap of both initiation time and execution time when the mild stutterers (M) are compared with their controls (C). Each speaker is represented twice in this figure, once for the immediate response condition and once for the delayed response condition. None of the differences between the fluent utterances of the mild stutterers and those of their controls was found to be statistically significant. The right side of Figure 1 indicates some overlap between severe stutterers and their controls in initiation times, but no overlap in execution time. Only two of the severe stutterers were judged to produce fluent utterances, but they were both slower than their controls in the execution of the number series whether the response was delayed or immediate.

Finger Task

Stutterers, on the average, were found to be about 14% slower than controls in the finger task. Table 3 summarizes the means and standard deviations of measures taken for each subject. Differences between groups were not found to be significant, however, with t tests applied to the initiation times in delayed and immediate conditions or to execution times in the immediate condition. There was too much overlap--some of the stutterers were quite fast, while some of the controls were relatively slow. A significant difference was found, however, between the groups in the mean times taken to execute the series in the delayed condition ($t(14) = 2.34$, $p < .025$). Again, when the stutterers were grouped according to severity, the severe stutterers accounted for differences found in the pooled data. Severe stutterers were significantly slower than their controls in the times taken to execute the series--in both immediate execution ($t(6) = 2.85$, $p < .025$) and delayed execution ($t(6) = 4.64$, $p < .005$) conditions. The severe stutterers were also significantly slower than their matched controls in initiation time in the immediate response condition ($t(6) = 2.23$, $p < .05$) but not when the signal to respond was delayed.

Figure 2 illustrates the extent of the overlap of mild stutterers and their controls in contrast with the separation of the data points for the severe stutterers and their controls, especially for execution time. No significant difference was found between the mild stutterers and their controls in finger counting. An analysis of the errors excluded from the data revealed that although only one of the control subjects and two of the stutterers made no errors, the errors (missed finger contacts and number reversals) averaged 3.7 for the controls, and 2.7 for stutterers for the list of 20 number series. A one error difference did not seem sufficient to account for the differences in speed between the groups.

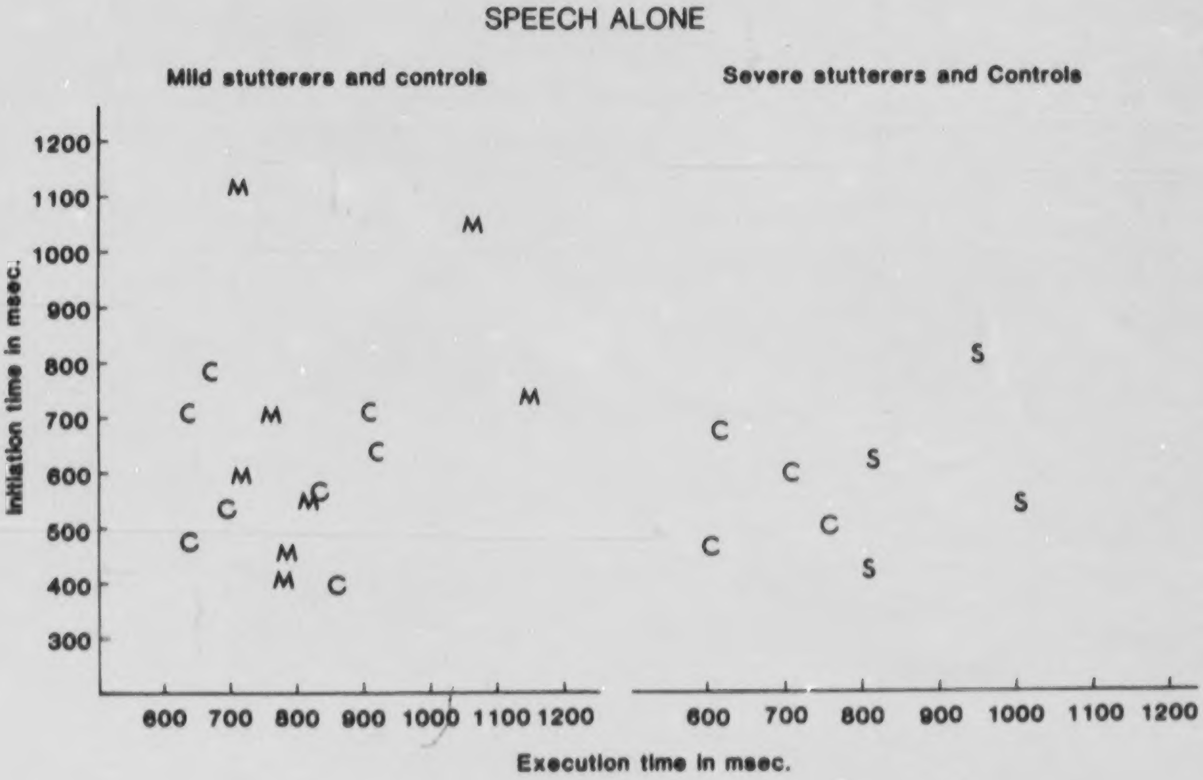


Figure 1. Mean initiation times plotted by mean execution times during the speech counting task for mild stutterers (M) with their matched controls (C) and severe stutterers (S) with their matched controls (C).

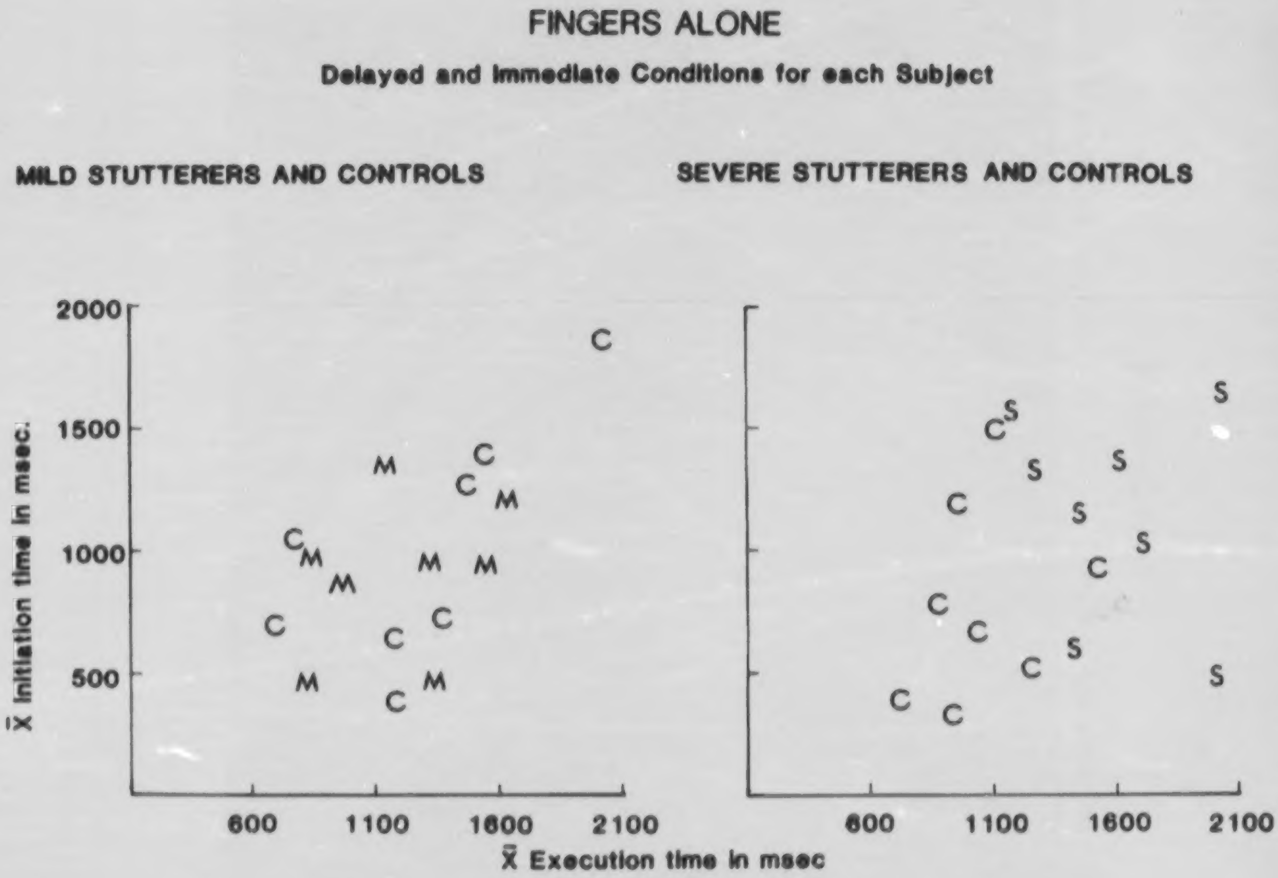


Figure 2. Mean initiation times plotted by mean execution times during the finger counting task for mild stutterers (M) with their matched controls (C) and severe stutterers (S) with their matched controls (C).

FINGER COUNTING

$\bar{X}$ and (SD) in msec.

Experimental Group

Subjects	Delayed Initiation	Immediate Initiation	Delayed Execution	Immediate Execution
1 S	497 (231)	1038 (138)	2014 (794)	1713 (210)
2 S	617 (105)	1134 (192)	1439 (309)	1482 (317)
3 S	1373 (588)	1624 (986)	1607 (740)	2018 (621)
4 M	948 (310)	1203 (898)	1562 (503)	1617 (1113)
5 S	1313 (431)	1566 (562)	1269 (263)	1163 (199)
6 M	476 (239)	986 (360)	1335 (430)	1324 (321)
7 M	845 (341)	1350 (287)	982 (112)	1144 (322)
8 M	452 (222)	986 (189)	815 (301)	845 (145)
Grand $\bar{X}$ (SD)	815 (347)	1236 (237)	1378 (350)	1413 (348)

Control Group

1 C	518 (567)	931 (247)	1246 (310)	1527 (683)
2 C	335 (56)	668 (199)	830 (90)	1035 (359)
3 C	1188 (585)	1497 (364)	959 (385)	1115 (497)
4 C	1387 (618)	1852 (447)	1553 (362)	2057 (490)
5 C	381 (110)	784 (230)	729 (39)	856 (206)
6 C	385 (42)	638 (98)	1167 (183)	1175 (94)
7 C	699 (324)	1026 (192)	696 (157)	781 (206)
8 C	718 (421)	1288 (224)	1384 (442)	1480 (348)
Grand $\bar{X}$ (SD)	701 (367)	1086 (402)	1071 (295)	1253 (391)

Table 3. Means and standard deviations of finger contact intervals in milliseconds.

Speech and Finger Counting Compared

The manual task was about 60% slower, on the average, than the speech task for both stutterers and controls (Table 4). There was more variability in timing for the finger counting than there was for speech counting for both groups. The advantage of knowing ahead (delayed condition) was evident for both groups in the initiation time required for both tasks. This advantage did not extend into the execution of the last three digits during speech as it did for the finger counting task.

DISCUSSION

An interesting finding of this study is the lack of significant differences between mild stutterers and their controls, in contrast with the significant differences found when severe stutterers were compared with their controls. This contrast is obscured when stutterers are pooled regardless of severity. Few studies have explored the timing of fluent utterances according to severity of stuttering. There were no stutterers in the present study that were judged moderate; they were either mild or severe. The stutterers who participated in the present study also served as subjects for another study of laryngeal reaction time (Alfonso, Watson, & Russo, Note 2). They found significant differences between the severe stutterers and controls for 13 different foreperiods (intervals between warning signal and cue to say 'ah'), but no significant differences between the mild stutterers and controls were found for 12 of the intervals. At the shortest foreperiod (100 msec), however, for mild stutterers the latency of voice onset was significantly different from controls. Another study that classified stutterers instead of pooling them compared elementary school children who stuttered and who also exhibited other mild to moderate articulation or language disorders with children who simply stuttered (Cullinan & Springer, 1980). The children with additional disorders took significantly longer than nonstutterers to initiate and to terminate voicing, while children who simply stuttered were not significantly slower than the controls. These studies, along with the present study, suggest that we may be losing important information by pooling data for stutterers. Specifically, there may be stutterers who have a more generalized motor coordination problem underlying their dysfluencies, and other stutterers for whom this deficit is confined to speech. When fluent, mild stutterers may be more similar to normal speakers than they are to severe stutterers.

One cannot compare this study to most previous reaction time studies, because the tasks here involved serial ordering of speech instead of simpler phonatory responses. Previous reaction time studies, cited in the introduction, required speakers to utter a single speech sound or a known word and sometimes to press a button or key.

A comparison of this study with other studies of manual versus oral timing is also difficult due to procedural differences. Other studies have required a simple flexor response of key pressing, an anticipated response, while this study required a serially-ordered response with coordination of many muscle groups and, in the immediate condition, the exact response could not be anticipated. Considering the initiation times alone, the present study would support those studies that found no significant difference between

SPEECH AND FINGER COUNTING

$\bar{X}$ and (SD) in msec.

	Delayed Initiation	Immediate Initiation	Delayed Execution	Immediate Execution
Experimental Group:				
Fingers	815 (347)	1236 (237)	1378 (350)	1413 (348)
Speech (fluent)	523 (115)	802 (207)	876 (154)	856 (119)
Control Group:				
Fingers	701 (367)	1086 (402)	1071 (295)	1253 (391)
Speech	470 (83)	648 (75)	725 (100)	713 (94)

Table 4. Means and standard deviations of intervals in speech and finger tasks compared.

stutterers on the average and their controls in the manual task (Reich et al., 1981), but when the severe stutterers were separated from the others, a significant difference was found in initiation times when the response was required to be immediate. Execution time has not been explored in other studies, but the present finding of significantly longer execution times for severe stutterers suggests that some stutterers need more time to coordinate serially-ordered events, regardless of whether they involve speech or hand coordination. Before offering possible explanations for these results, a caveat is in order. A separate aspect of this experiment required that the subject perform the speech task first to increase the possibility that stuttering samples would be obtained in addition to the fluent tokens. It is possible that the state of excitability for the speech task carried over into the finger counting task. Thus, we must view our conclusions with caution. We are left with at least three possibilities: 1) a radiation effect: discoordination of fine motor control in severe stutterers that includes not only speech muscles but hand muscles, 2) a generalized arousal effect carried over from performing the speech task before the finger task, and/or 3) a speech mediation effect, in which the finger task took longer to execute not due to any problem in hand coordination but due to the possibility that subjects were "speaking to themselves" as they counted on their fingers. Further research is needed to test these possibilities.

On the question of whether knowing the expected response one second ahead of the response signal extends the advantage given to initiation into the execution of the rest of the series, the interesting finding was that the utterances of normal speakers and the fluent tokens of stutterers were similar, in contrast with stutterers' dysfluent tokens. All subjects took less time to initiate the task in the delayed-response conditions, whether finger or speech counting, but the fluent tokens of stutterers were like their controls in that this advantage failed to extend through the execution of the last three digits of the spoken series. When the series was stuttered, however, the stuttering was prolonged further in both initiation and execution phases when the response signal was immediate rather than when delayed. The obvious cases of "jumping the gun" in the delayed condition were removed from the analysis, but it remains possible that the measured times of delayed initiation may be artificially shortened by some anticipation by both groups. The effect is probably spread across groups, however, as the ratios between delayed and immediate conditions of initiation are similar for both fluent stutterers (1:1.5) and controls (1:1.4), with the initiation demanded by immediate response taking about half again as long as under the delayed condition.

For the speech task, this study has gone one level further than other studies in delineation of "fluent" utterances of stutterers. To qualify as fluent, the utterances were perceptually fluent to an observer, by both eye and ear, and, in addition, were "physiologically fluent" by examination of the movement indices as inferred from the lower lip trace, the laryngeal impedance changes, and the respiratory traces. Any abnormal perturbation in the traces was considered as evidence that the utterances fell outside the boundaries of fluency. All such utterances were discarded from the fluent sample.

Since stutterers evidence most of their dysfluencies during the initiation of phrases rather than within phrases, it was interesting and surprising that initiation times for the fluent utterances were not significantly longer than controls, while execution times were significantly longer. Initiation of sequential speech demanded by the present study required much more than initiation of voice. It demanded the visual perception of the series to be executed, pre-movement motor readiness including excitation of the motoneuron nets to be involved, and finally, the specific neuromotor and myomotor events leading to the movements recorded. It included production of the voiceless consonant and the motor adjustments preparatory to voicing the first number of each series. Stuttering did occur on the first digit for 86% of the stuttered utterances, whereas the incidence dropped to 42% for the second, 46% for the third, and 26% for the last digit. When the tokens of stutterers were judged to be fluent, however, the times taken to initiate the response were not significantly longer even though the utterances were executed more slowly. These results lend support to the notion that it may take no more time for a stutterer to prepare for a fluent utterance than it does for a nonstutterer; it is only when the preparation is faulty that the stutterers block initiation of the speech. Faulty preparation might involve either the generation of an insufficient or excessive degree of excitability of appropriate neural networks. (Evidence for preparatory adjustments preceding movement and the difficulties in specifying them are reviewed by Requin, 1980.)

The principle of selective potentiation is thought to play a part in motor coordination; that is, the system increases the potential for certain neural activity while reducing the potential for activity in other neural circuits (Gallistel, 1980). In discoordinated motor acts, there may be a failure to achieve a state of arousal that is optimal for the task, and neural nets that serve a particular group of muscles may be overexcited while other groups may be underexcited (see Zimmermann, 1980b). The state of equilibrium among cooperating units and agonist-antagonist units that allows for reciprocal inhibition may not be achieved (Freeman & Ushijima, 1978). On the other hand, if stutterers achieve a balanced pre-movement set, they may be fluent and the set will take no more time than it would for nonstutterers. If their settings are faulty, one would expect the initiation of a coordinated act to be the most difficult part; once started it would be easier to complete.

Why, then, were the severe stutterers slower than their controls in the execution of the sequences? Was slowing the response the price that they paid for fluent performance? In order to maintain relative fluency, are there changes in the temporal organization of the mechanisms coordinating for speech? The author is currently analyzing the differences in coordination among the respiratory, laryngeal, and supralaryngeal movements recorded during stuttered utterances, perceptually fluent utterances, and control utterances. Differences in coordination patterns may be found to relate to the slowing of execution, even when "fluent."

REFERENCE NOTES

1. Luper, H. L., & Cross, D. E. Relation between finger reaction time and voice reaction time in stuttering and nonstuttering children and adults. Paper presented at a convention of the American Speech and Hearing Association, San Francisco, 1978.

2. Alfonso, P. J., Watson, B. C., & Russo, A. *Variable foreperiod effects on laryngeal reaction time in stutterers and normal speakers*. Paper presented at a convention of the American Speech, Language, and Hearing Association, Los Angeles, November 1981.

REFERENCES

Adams, M. R., & Hayden, P. The ability of stutterers and nonstutterers to initiate and terminate phonation during production of an isolated vowel. *Journal of Speech and Hearing Research*, 1976, *19*, 290-296.

Bloodstein, O. *A handbook on stuttering*. Chicago: National Easter Seal Society, 1975.

Cullinan, W. L., & Springer, M. T. Voice initiation and termination times in stuttering and non-stuttering children. *Journal of Speech and Hearing Research*, 1980, *23*, 344-360.

Freeman, F. J., & Ushijima, T. Laryngeal muscle activity during stuttering. *Journal of Speech and Hearing Research*, 1978, *21*, 538-562.

Gallistel, C. R. *The organization of action: A new synthesis*. Hillsdale, N.J.: Lawrence Erlbaum, 1980.

McFarlane, S. C., & Prins, D. Neural response time of stutterers and nonstutterers in selected oral motor tasks. *Journal of Speech and Hearing Research*, 1978, *21*, 768-778.

Ostry, D. J. Execution-time movement control. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North-Holland, 1980, 457-468.

Prosek, R. A., Montgomery, A. A., Walden, B. E., & Schwartz, D. M. Reaction-time measures of stutterers and nonstutterers. *Journal of Fluency Disorders*, 1979, *4*, 269-278.

Reich, A., Till, J., & Goldsmith, H. Laryngeal and manual reaction times of stuttering and nonstuttering adults. *Journal of Speech and Hearing Research*, 1981, *24*, 192-196.

Requin, J. Toward a psychobiology of preparation for action. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior*. Amsterdam: North Holland, 1980, 373-398.

Riley, G. D. A stuttering severity instrument for children and adults. *Journal of Speech and Hearing Disorders*, 1972, *37*, 314-322.

Ryan, B. *Programmed therapy for stuttering in children and adults*. Springfield, Ill.: Charles C. Thomas, 1974.

Starkweather, C. W., Hirschman, P., & Tannenbaum, R. S. Latency of vocalization onset: Stutterers versus nonstutterers. *Journal of Speech and Hearing Research*, 1976, *19*, 481-492.

Starkweather, C. W., & Myers, M. Duration of subsegments within the intervocalic interval in stutterers and nonstutterers. *Journal of Fluency Disorders*, 1979, *4*, 205-214.

Zimmermann, G. Articulatory dynamics of fluent utterances of stutterers and nonstutterers. *Journal of Speech and Hearing Research*, 1980, *23*, 95-107. (a)

Zimmermann, G. Stuttering: A disorder of movement. *Journal of Speech and Hearing Research*, 1980, *23*, 122-136. (b)

TRADING RELATIONS IN THE PERCEPTION OF SPEECH BY FIVE-YEAR-OLD CHILDREN

Rick C. Robson+, Barbara A. Morrongiello+,++, Catherine T. Best+++, and Rachel K. Clifton+

Abstract. Five-year-old children were tested for perceptual trading relations between a temporal cue (silence duration) and a spectral cue (F_1 onset frequency) for the "say"-"stay distinction." Identification functions were obtained for two synthetic "say"-"stay" continua, each containing systematic variations in the amount of silence following the /s/ noise. In one continuum, the vocalic portion had a lower F1 onset than in the other continuum. Children showed a smaller trading relation than has been found with adults. They did not differ from adults, however, in their perception of an "ay"-"day" continuum formed by varying F1 onset frequency only. The results of a discrimination task in which the two acoustic cues were made to "cooperate" or "conflict" phonetically supported the notion of perceptual equivalence of the temporal and spectral cues along a single phonetic dimension. The results indicate that young children, like adults, perceptually integrate multiple cues to a speech contrast in a phonetically relevant manner, but that they may not give the same perceptual weights to the various cues as do adults.

In the developmental literature on speech perception, there are several reports that children differ from adults in their responses to variations in single acoustic cues for phonetic contrasts. Zlatin and Koenigsknecht (1975), studying the perception of the stop consonant voicing contrast in two-year-old, six-year-old, and adult listeners, found that the magnitude of voice-onset-time (VOT) difference necessary for distinguishing between prevocalic stop cognates decreased as a function of age. Simon and Fourcin (1978) varied both VOT and first-formant (F1) transition steepness in an investigation of two- to fourteen-year-old English and French children's perception of voicing oppositions. The authors were particularly interested in studying French speakers' perception of voicing, since the VOT boundary differs from English and the F1 transition is a more salient cue in French than in English. Their

---

+University of Massachusetts, Amherst, MA.
++University of Toronto/Erindale College, Mississauga, Ontario CANADA.
+++Also Neuroscience and Education Department, Teachers College, Columbia University, New York, NY.

Acknowledgment. The authors wish to thank the parents and children who participated in this research; their interest, patience, and good humor made this research possible and contributed to making the conduct of this study a most enjoyable experience. This research was supported by NIH grant HD06753-06 awarded to Rachel Clifton, by NICHD Grant HD01994 awarded to Haskins Laboratories, and by postdoctoral fellowship grant NS5085 awarded to Catherine Best.

results revealed a linear improvement in labeling accuracy with age for children of both language environments, with an adult-like categorical pattern occurring at five to six years for the English and seven to eight years for the French listeners. Moreover, English-speaking children showed no evidence of utilizing the F1 transition cue before about five years of age. The phoneme boundary between voiced and voiceless percepts also showed a systematic shift until 11 or 12 years of age when it reached a value corresponding to adult performance.

While these differences between children's and adults' phonetic perception, as based on single acoustic cues, are interesting, evidence is accumulating in the adult speech perception literature, that multiple acoustic cues often interact to specify a single phonetic contrast. For example, voicing distinctions for initial stop consonants can be cued by changes in VOT, F1 onset frequency, F0 contour, or aspiration energy (Haggard, Ambler, & Callow, 1970; Lisker, 1975; Lisker, Liberman, Erickson, Dechovitz, & Mandler, 1977; Repp, 1979); each of these acoustic properties is a consequence of the laryngeal timing variations underlying the production of stop voicing (Abramson & Lisker, 1965). Multiple acoustic correlates of articulatory contrasts have also been found to serve as cues for the perception of place of articulation (Dorman, Studdert-Kennedy, & Raphael, 1977; Harris, Hoffman, Liberman, Delattre, & Cooper, 1958) and manner of articulation (Dorman, Raphael, & Liberman, 1979; Miller & Liberman, 1979; Repp, Liberman, Eccardt, & Pesetsky, 1978).

Whenever several distinct acoustic cues provide listeners with functionally equivalent information about a single phonetic category contrast, then perceptual "trading relations" can be demonstrated. That is, strengthening the value of one cue can offset the weakening of another in listeners' perception of the specified phonetic contrast. Such trading relations have been found for voicing (e.g., Summerfield & Haggard, 1977), place (e.g., Bailey & Summerfield, 1980), and manner of articulation (e.g., Dorman, Raphael, & Isenberg, 1980) distinctions.

In a recent series of experiments, we examined the perceptual equivalence of acoustic cues in adults' perception of speech and related nonspeech sounds (Best, Morrongiello, & Robson, 1981). Using a "say"-"stay" (/sei/-/stei/) contrast, we systematically manipulated two acoustic cues that specify the presence or absence of the alveolar stop following the word-initial /s/: F1 onset frequency and the duration of the silent closure interval. The average trading relation obtained from listeners' identification performance was evident in a "say"-"stay" boundary shift of 24.6 msec (Experiment 1). In other words, in order to be perceived as "stay," a stimulus with a high F1 onset frequency (430 Hz) required approximately 25 msec additional silence between the /s/ and the vocalic portion than did a stimulus token having a low F1 onset frequency (230 Hz).

To provide a more stringent test of whether these two acoustic cues were truly equivalent in perception (cf. Fitch, Halwes, Erickson, & Liberman, 1980), discrimination performance was assessed for stimulus comparisons in which the parameter values for closure duration and F1 onset frequency were either "cooperating" (i.e., complementing one another phonetically) or "conflicting" (i.e., cancelling each other). Since the Cooperating Cues and the

Conflicting Cues conditions differed only in the combination of cue values but not in the magnitudes of differences on each cue dimension, performance in the two conditions should have been equal if listeners discriminated the stimuli by their auditory properties alone. In contrast, listeners performed near chance in the Conflicting Cues condition but at a much higher level in the Cooperating Cues condition. Thus the results supported the hypothesis that the two acoustic cues provide perceptually indistinguishable ("perceptually equivalent") information along a single phonetic dimension.

In the present research we extended our investigation to children's speech perception. By using the same stimuli as in the Best et al. (1981) study, we sought to determine whether children show a phonetic trading relation and perceptual equivalence of acoustic cues to the /sei/-/stei/ contrast in the same manner as adults do. Children five years of age were tested, since this was the age at which Simon and Fourcin (1978) claimed to first find evidence of perceptual use of F1 transition distinctions in perception of stop voicing contrasts. Children's identification performance was assessed by using a standard forced-choice procedure. However, Wolf (1973) reported that five- and seven-year-old children have difficulty with the ABX discrimination procedure, and pilot testing in our laboratory confirmed this observation. Consequently, discrimination data were obtained using a 2IAX paired-comparison procedure, in which children judged the pair members as being the "same" or "not the same" (Wolf, 1973).

Since there was some evidence to indicate developmental changes in perception of VOT (Bernstein, 1979; Simon & Fourcin, 1978; Zlatin & Koenigsknecht, 1975) and in the location and stability of various phoneme boundaries in perception and production (Kewley-Port & Preston, 1974; Strange & Broen, 1981; Zlatin & Koenigsknecht, 1976), we expected that children might differ from adults in performance on our multiply-cued stimulus continuum, which involved variations in F1 onset frequency and in a temporal cue (as in VOT). The developmental literature, however, did not support a particular hypothesis as to the nature of these potential age-related differences (e.g., better utilization of the spectral than of the temporal cue or vice versa), although evidence that young children are less sensitive than adults to small differences in formant frequency information (Eguchi, 1976) suggested that five-year-olds might be less responsive to F_1 onset manipulations than adults.

Although Simon and Fourcin (1978) claim that English-speaking children begin to make perceptual use of a temporal cue to stop voicing earlier than they make use of a spectral cue, there are some methodological problems with their study.[1] Insofar as Simon and Fourcin's findings generalize to children's perceptual integration of slightly different temporal and spectral cues for a different phonemic contrast, they suggest that the children in our study might attend more to the temporal than the spectral cue and hence show a smaller trading relation than the adults in Best et al. (1981). However, even if the children do show a reduced trading relation, there is no indication in the developmental literature as to whether a discrimination test would reveal the same perceptual equivalence of the two cues along a single phonetic dimension as was found in adults. The present study was undertaken to assess whether 5-year-olds make perceptual use of multiple cues for a single phonemic contrast in a manner that indicates attention to phonetic information, as adults do. Alternatively, if children attend primarily to the acoustic properties of the stimuli, then one would expect that they would fail to integrate perceptually the temporal and spectral cues as information about a unified phonetic category. In that case, they would hear the auditory differences between differently-cued stimuli even within a phonetic category, and would thereby discriminate the Conflicting Cues contrasts as well as they discriminate the Cooperating Cues contrasts. Although this second possibility was less likely on the basis of the adult findings, it could not be dismissed *a priori* because no studies of trading relations in children existed in the literature.

METHOD

Subjects

Eight children (3 male, 5 female) approximately five years old at the onset of testing (mean age, 60.4 months; range; 57.3-64.9 months) participated in the present experiment. An average of 3 1/2 months elapsed between the first and final testing sessions. Children were reported by parents to have normal hearing and did not have colds, ear, or throat disturbances on test days. The data from two additional children were excluded from the final analysis because of incomplete test sessions. Parents were paid $3.00 for transportation costs, and children selected a prize for each day of participation.

Stimuli

Two sets of synthetic stimuli were used. They were based upon two 290-msec, three-formant syllables created on the Haskins parallel-resonance synthesizer (see Figure 1), as stylized versions of the vocalic portions of natural utterances of "say" and "stay" produced by a male speaker. They differed from one another only in F1 onset frequency (230 Hz vs. 430 Hz). The syllables were identical in formant amplitudes and overall amplitude envelopes, in F2 and F3, and in the F1 steady-state frequency (611 Hz) beyond the initial 40-msec transition difference (see Best et al., 1981, for complete stimulus descriptions).

One set of stimuli was an "ay-day" continuum[2] spanning 14 different syllables. It was created by varying the F1 onset frequency in approximately 33 Hz steps between 160 Hz and 611 Hz, and included the 230 Hz and 430 Hz F1 onset syllables described above. In a previous identification test using the "ay-day" continuum (Best et al., 1981), adults identified the 230-Hz syllable as "day" 100% of the time. This syllable will hereafter be referred to as the "strong day," abbreviated D. In contrast, adults identified the 430-Hz syllable as "day" only approximately 50% of the time; therefore, it will be called the "weak day," abbreviated d. To test whether the two test syllables would also differ in children's perception, a stimulus tape was constructed for obtaining the children's identification functions on the "ay-day" continuum. The tape contained ten presentations of each of the 14 syllables in a randomized sequence. Within each block, the intertrial interval was 4 seconds.

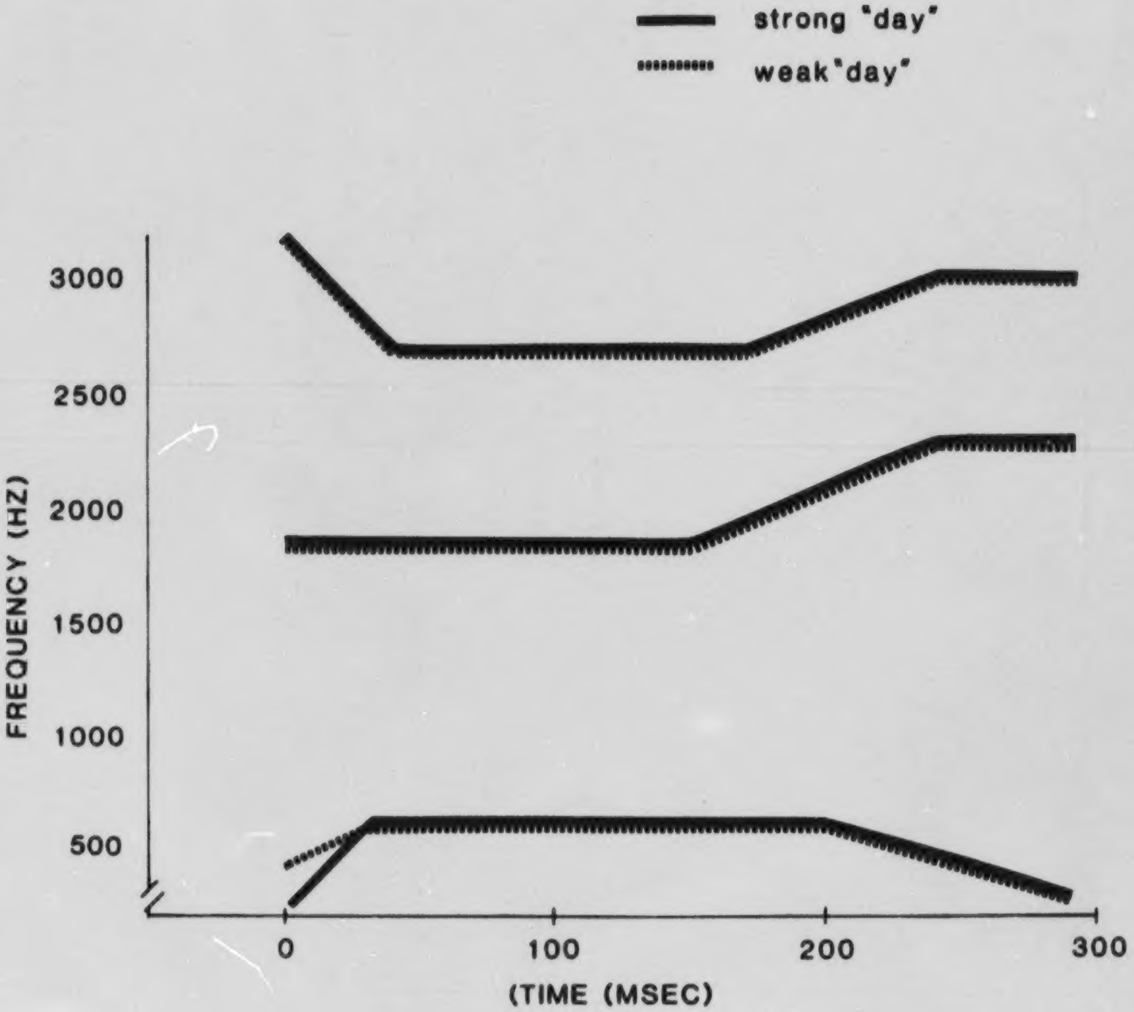


Figure 1. Schematic diagram of F_1, F_2, and F_3 frequencies for synthetic "weak day" and "strong day."

Table 1

Stimulus Pairings for the Four Discrimination Conditions

One Cue[a]	Cooperating Cues	Conflicting Cues	Physically Same[b]
s - 24 - D vs. s - 24 - d	s - 32 - D vs. s - 8 - d	s - 8 - D vs. s - 32 - d	s - 0 - d/D
s - 32 - D vs. s - 32 - d	s - 40 - D vs. s - 16 - d	s - 16 - D vs. s - 40 - d	s - 8 - d/D
s - 40 - D vs. s - 40 - d	s - 48 - D vs. s - 24 - d	s - 24 - D vs. s - 48 - d	s - 96 - d/D
	s - 56 - D vs. s - 32 - d	s - 32 - D vs. s - 56 - d	s - 104 - d/D

[a] 's' stands for the /s/ portion of the syllable; the subsequent number is the number of msec silence between the /s/ and the vocalic portion of the syllable; 'D' stands for the "strong day" syllable and 'd' stands for the "weak day" syllable.

[b] Because the members of a pair here are physically the same, only one member of each pair type is shown; d/D indicates there was one "weak day" pair and one "strong day" pair.

The second set of stimuli consisted of two different "say-stay" continua, constructed by preceding the D and d syllables with a natural 120-msec /s/ noise derived from a male speaker's utterance of "say" (see Experiment 2 of Best et al., 1981). The /s/ and the synthetic syllable were separated by silent intervals ranging from 0 to 104 msec, in 8-msec increments. Thus, each continuum comprised 14 tokens.

Two stimulus tapes were constructed. The first tape was designed to obtain children's identification functions. This tape consisted of 20 blocks of 14 single-item trials each. Every two successive blocks comprised a randomized sequence of all 14 tokens from each of the two continua, for a total of 10 repetitions per token. Within each block, the intertrial interval was 4 seconds.

The second tape constructed from the "say-stay" stimuli was used to test discrimination. A 2IAX discrimination task ("same"-"not same") was employed. This test included four types of stimulus pairings for discrimination judgments: Physically Same, One Cue, Conflicting Cues, and Cooperating Cues (see Table 1). There were 8 different Physically Same pairs, four from each of the two "say"-"stay" continua. These four pairs were based on the two extreme endpoints of each continuum, which were clear instances of "say" or "stay." There were also three different pairs for the One Cue comparisons. Within each One Cue pair, the tokens were identical in silent gap duration, but differed in the spectral cue (d vs. D). These three pairs were selected so that the silent gap durations spanned the adult "say"-"stay" boundaries (lower panel of Figure 2), as determined by Experiment 1 of Best et al. (1981). In both the Cooperating and the Conflicting Cues comparisons, also referred to as the Two Cue comparison types, members of each discrimination pair differed on both the spectral and the temporal dimension. In the Cooperating Cues comparisons, the D member of the pair had a 24-msec longer silent gap duration than the d member (as in Experiment 1 of Best et al.); thus the temporal and spectral cue values for each pair member "cooperated" in that they both favored the same phonetic category. In the Conflicting Cues comparisons, the D member of a pair had a 24-msec shorter silent gap duration than the d member. Here, the value of the temporal cue was designed to cancel the phonetic effect of the spectral cue for each pair member. In both the Two Cue comparison types, a 24-msec difference in silent gap duration was used because this was the magnitude of the trading relation shown by adults for identifications of the two stimulus continua (Experiment 1, Best et al., 1981). There were four different pairs in each of the Two Cue comparison types, selected so as to span the "say"-"stay" boundaries for adults.

The discrimination tape contained 240 trials organized into 16 blocks of 15 trials each. The 19 different stimulus pairs (eight Physically Same, three One Cue, four Cooperating Cues, four Conflicting Cues) were randomly sequenced within each successive pair of blocks. Within each pair of blocks, each of the "not same" pairs (One Cue, Cooperating Cues, Conflicting Cues comparisons) was presented twice, whereas each of the Physically Same pairs was presented once. Thus, 16 judgments were obtained for each of the "not same" pairs, and eight for each of the Physically Same pairs. The interstimulus interval within each pair was 1 second, and the intertrial interval between successive pairs was 4 seconds.

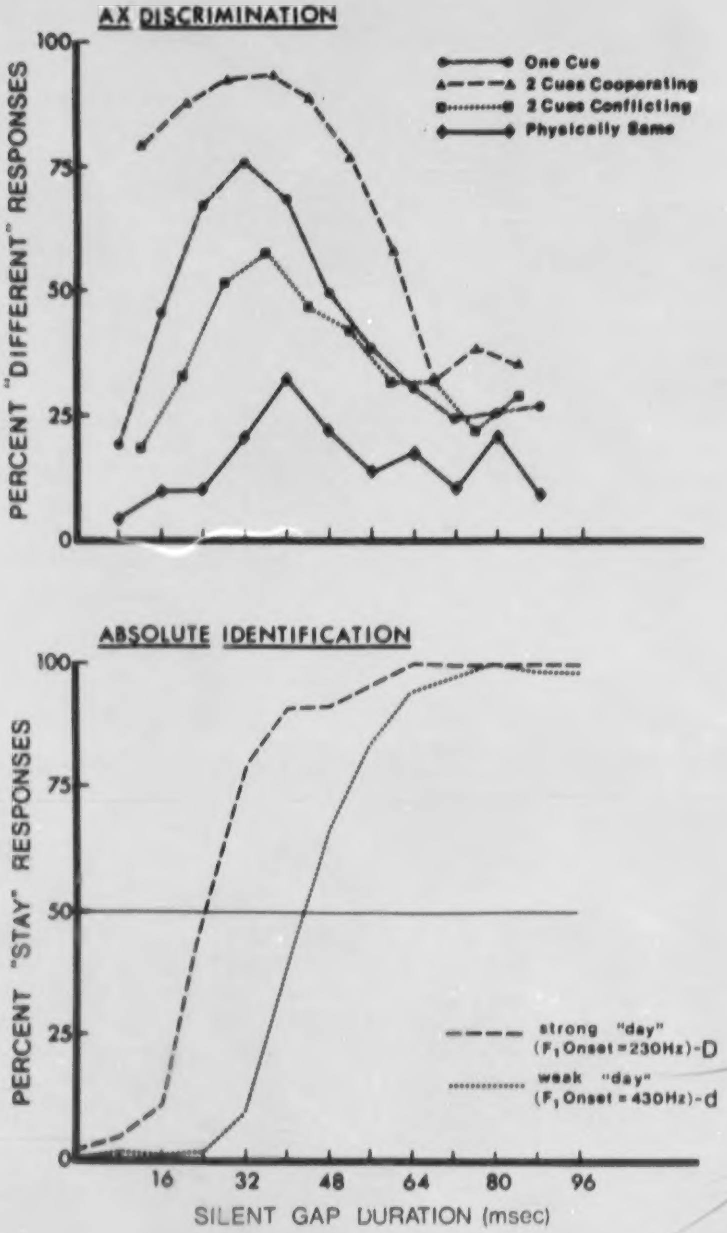


Figure 2. Obtained functions for the three-way 2IAX discrimination test ("same"-"different"; upper panel) and the forced choice identification test on the two "say-stay" stimulus continua (lower panel) for the adults tested in Experiment 2 of Best, Morrongiello, and Robson, Perception & Psychophysics, 1981, 29, 191-211 (Reprinted with publisher's permission).

Apparatus and Procedure

Each child participated in five 50-minute sessions conducted within a few weeks of one another. The first and second halves of the "say-stay" identification test were given in sessions 1 and 3, and the two halves of the 2IAX "say-stay" discrimination test were given in sessions 2 and 4. In session 5, the randomized forced-choice "ay-day" identification test was given. Testing was conducted in a sound-attenuated room with the parent and Experimenter 1 present. The stimuli were played on a Revox reel-to-reel tape recorder running at 7.5 ips at a Sound Pressure Level of 60 dB re .0002 dynes cm^2 (calibrated using the A scale of a General Radio sound level meter) over loudspeakers (Acoustic Research, #AR-7) located approximately 1 m to the child's left and right, at a 90-degree angle to the child's midline.

Upon entering the testing room, children were given five minutes to become accustomed to the new situation. During this time Experimenter 1 encouraged the child to play with two small mechanical robots. Once rapport had been established, the child was told that a big robot in the adjacent equipment room was learning how to speak and that she/he could help the robot learn to talk better. Most children were enthusiastic about participating. After showing a child a robot that had been constructed around the tape recorder and having her/him listen and repeat the words that the robot said (i.e., taped versions of clear endpoint "say" and "stay"), children were taught to use a two-button box in the testing room to indicate their responses "to the robot" in the equipment room. An Esterline-Angus event recorder in the equipment room recorded the child's responses on the two-button box. Throughout the test session, Experimenter 2 tallied the child's responses directly from the Esterline-Angus recorder and indicated interblock intervals on the permanent paper record. After the test session, the tally completed by Experimenter 2 was checked by a naive observer against the permanent paper tape record.

During the "say-stay" identification tests, the child pressed either of two horizontally-adjacent buttons on the button-box to indicate whether "say" or "stay" was heard on each presentation. A picture adjacent to each button was a continuous reminder of which button was for "say" (i.e., a picture of a woman talking and the word "say" printed) and which button was for "stay" (i.e., a picture of a woman motioning for her dog to stay and the word "stay" printed). For the "ay-day" test, the pictures used were of a large letter "A" for "ay" and a sun rising over the horizon for "day." The right-left button designation for each word was randomized across test sessions and children.

During the 2IAX discrimination test, two strips of colored tape were substituted for each picture on each button box. For one button the two colors were the "same" (both red) and for the other button the two colors were "not the same" (red and green). During the 2IAX discrimination test the children were instructed to listen to each pair of words and press a button to indicate whether the pair members were exactly the "same" or "not the same." Again, the right-left button designation was randomized across sessions and children.

On each day of testing the child was reminded of how to use the response box, and was given a block of practice trials to insure that she/he understood the task and could work through an entire block of trials without difficulty. Experimenter 1 remained with the child throughout each test session and provided verbal encouragement and support, as necessary. In addition, throughout the testing sessions two low-watt blue spot-lights provided the child with intermittent feedback, which proved to be particularly effective in motivating the child to perform the task and continue to listen closely. The lights were positioned approximately 1 m in front of the child. On one light a happy face signaled that the child's previous response had been correct. A sad face on the other light indicated an incorrect response. Experimenter 2 controlled the operation of these lights according to the correctness of the child's responses on a sample of trials. During the "say-stay" identification sessions, one of each of the endpoint stimuli for the two continua was randomly selected during the course of two trial blocks for reinforcement. During the discrimination sessions, one of each of the four types of trials was selected and for the "ay-day" identification series, one of each of the endpoint stimuli for the continuum received reinforcement.

Between trial blocks in all five sessions, children were allowed to select colored stars that they pasted on a personalized game board. On successive blocks they selected an increasing number of stars and after the last trial block they were allowed to select a prize. For most of the children the time during which they selected and pasted stars was sufficient to serve as a rest interval. However, when necessary for maintaining the child's motivation for the test sessions, this inter-block interval was lengthened and the child was allowed to engage in another play activity for a few minutes.

RESULTS

Identification: "say-stay"

The category boundary between "say" and "stay" was defined as that silent interval at which there were 50% "stay" responses. There were no significant test block effects (session 1 vs. 3) in the children's identification responses. As can be seen in Figure 3, the mean category boundary for the D continuum was at 26.4 msec (Range: 16.0-32.0 msec). In contrast, the mean category boundary for the d continuum was at 37.5 msec (Range: 33.6-43.3 msec). This average difference in category boundaries of 11.1 msec (Range: 5.9-17.6 msec) was highly significant ($t_7 = 8.5$, $p < .001$). In fact, there was no overlap whatsoever in the distribution of category boundaries for the D and d continua.

These results support previous findings, obtained with adults, of a trading relation between spectral and temporal acoustic cues in the perception of stop consonants. In children, "weak day" stimulus tokens required approximately 11.1 msec more silence after /s/ to be heard as "stay" than did "strong day" stimulus tokens (see Figure 3). The magnitude of this trading relation differs between children and adults ($t_{20} = 5.3$, $p < .001$).[3] This difference between children and adults is due exclusively to a difference in their identification of stimulus tokens from the d continuum (compare Figure 3 to the bottom panel of Figure 2). For the d continuum, the mean 50% crossover point for adults in Experiment 2 of Best et al. (1981) was 43.8 msec, whereas

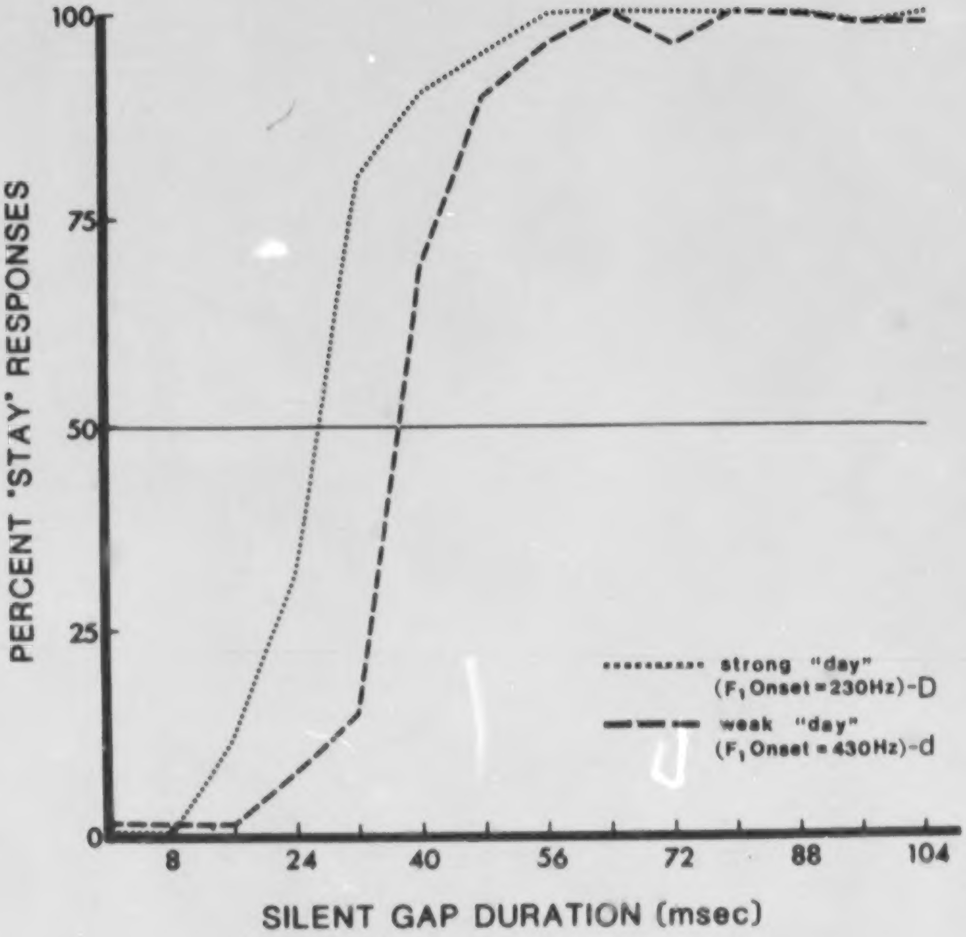


Figure 3. Identification functions of the children for the "strong day" and "weak day" stimulus continua.

that for children was 37.5 msec ($t_{20} = 2.2$, $p < .05$). For the D continuum the respective points were 25.3 msec and 26.4 msec ($t_{20} = .3$, n.s.).

Identification: "ay-day"

The results from the "ay-day" identification task may provide some insight into the basis for the difference between children and adults in the magnitude of the trading relation. Children were apparently not less sensitive than adults to the perceptual use of the F1 spectral cue for the alveolar stop, since as a group they did not differ significantly from adults in the location of the 50% crossover point for the "strong day" continuum. Rather, it was the 50% crossover point for the "weak day" continuum that differentiated the children and adults. One possibility is that children were more sensitive than adults to F1 onset spectral information, in the sense that for children a relatively high F1 onset supported perception of an alveolar stop, following /s/, more readily than it did for adults. Conversely, the children could be said to be less sensitive than adults to the spectral difference between the 230 Hz vs. 430 Hz F1 onsets. Since the "ay-day" identification task involved changes only in this spectral cue, it is useful for examining the possibility that the "weak day" vocalic syllable was perceived to be more "day"-like by children than by adults.[4]

The identification functions for the children, and for a sample of 18 adult listeners (Best et al., 1981), are shown in Figure 4. The 50% crossover point for the children did not differ significantly from that of the adults ($t\ 3 = .3$). The "ay-day" continuum contained the two vocalic syllable tokens used in generating the two "say-stay" continua ("weak day" continuum - 430 Hz F1 onset frequency; "strong day" continuum - 230 Hz F1 onset frequency). Children and adults did not differ in percent of "day" identification for either of these tokens: "strong day" token - adults 99%, children 100%; "weak day" token - adults 46%, children 54%. These results suggest that children's and adults' perception of the F1 onset spectral cue was not primarily responsible for the obtained difference in the size of the trading relation.

2IAX Discrimination Test

The discrimination data were compared with discrimination performance predicted from the identification data for the strong and weak "say-stay" continua. For a given discrimination comparison type, the probability of a "not same" response was computed in the following manner (see Best et al., 1981): p ("not same") = [p ("say" on first member of comparison) x p ("stay" on second member of comparison)] - [p ("stay" on first member) x p ("say" on second member)]. Since there were no significant effects involving blocks (i.e., testing session 2 vs. 4), only results totalled over blocks 1 and 2 will be reported. The results for Physically Same comparison types showed that there was no significant general response bias; the average observed proportion of "not same" responses was 4% and the average predicted proportion was 1%.

There are two aspects of discrimination performance that will be discussed: (1) observed vs. predicted performance for each discrimination type; and (2) the relative rank ordering of discrimination performance across discrimination types. With regard to the latter, it is important to remember

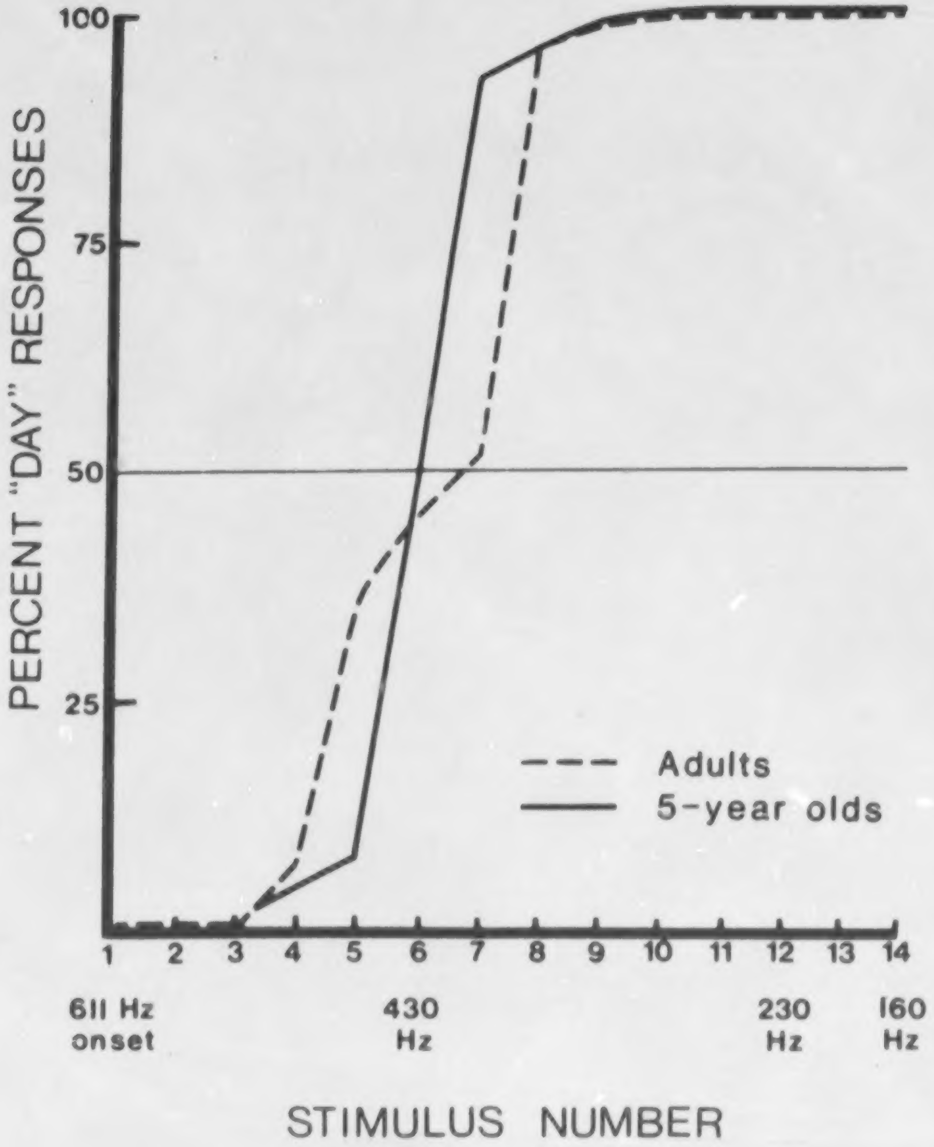


Figure 4. Children's and adults' identification functions for the "ay-day" stimulus series (stimulus numbers refer to steps of approximately 33 Hz in onset frequency of F_1; stimulus 6 is the "weak day" vocalic base of the continua used in the "say-stay" conditions, and stimulus 12 is the "strong day" stimulus base).

that in selecting stimulus pairs for the Conflicting Cues and Cooperating Cues discrimination types, a trading relation typical of adults was assumed (Experiment 1 of Best et al., 1981). Since the children in the study showed a significantly smaller trading relation than adults, however, the discrimination pairs used were not in fact appropriate for providing the most clear and dramatic contrast in the children's performance between the Conflicting and the Cooperating conditions. Specifically, instead of using a 24 msec silent gap difference between the members of Two Cue discrimination pairs, a difference of 11 msec would presumably have been more appropriate.

Nonetheless, the data can provide a test of the perceptual equivalence hypothesis if predicted and obtained discrimination performance were to vary in a similar manner as a function of discrimination condition, particularly if peak performance in the Cooperating Cues condition was still predicted to be higher than performance in the Conflicting Cues condition. To determine whether this was the case, an analysis of variance on predicted peak discrimination levels was performed for the Cooperating Cues, Conflicting Cues, and One Cue conditions. Peak performance was defined as performance on those comparisons in which the pair members straddled the "say"-"stay" boundary; that is, the second comparison for the One Cue condition, and the average of the second and third comparisons in each of the other two discrimination conditions. There was a significant difference among the conditions for the predicted discrimination data, $\underline{F}_{2,14} = 14.27$, $\underline{p} < .001$. Predicted performance was significantly higher for the Cooperating Cues than for the Conflicting Cues condition, $\underline{t}_7 = 4.93$, $\underline{p} < .01$, although the difference between the Conflicting Cues and the One Cue conditions was not significant. The observed vs. predicted scores for each test condition appear in Figure 5.

Analysis of variance on the observed performance levels also revealed significant differences among the conditions, $\underline{F}_{2,14} = 11.3$, $\underline{p} < .005$. The pattern of differences among the discrimination conditions conformed to predicted order, supporting the notion that children, like adults, perceived the diverse acoustic cues as equivalent information along a single phonetic dimension. Peak discrimination was significantly higher for the Cooperating Cues condition than the Conflicting Cues condition, $\underline{t}_7 = 3.6$, $\underline{p} < .01$. There was no significant difference between the Conflicting Cues and One Cue conditions.[5]

DISCUSSION

Investigation of trading relations among acoustic cues in phonetic perception can provide valuable insights into how information from diverse acoustic dimensions is integrated in the perception of speech. The present investigation examined children's integration of spectral and temporal cues for the perception of a stop consonant in an /s/ + stop cluster in syllable-initial position. Generally, to perceive the stop consonant children needed approximately 11 msec more silence to compensate for a weak spectral cue than when a strong spectral cue was present. This trading relation of 11 msec was significantly less than that obtained for a group of adult listeners tested with the same stimuli (Best et al., Experiment 2, 1981). Children and adults did not differ, however, in their perception of the "ay-day" continuum, which

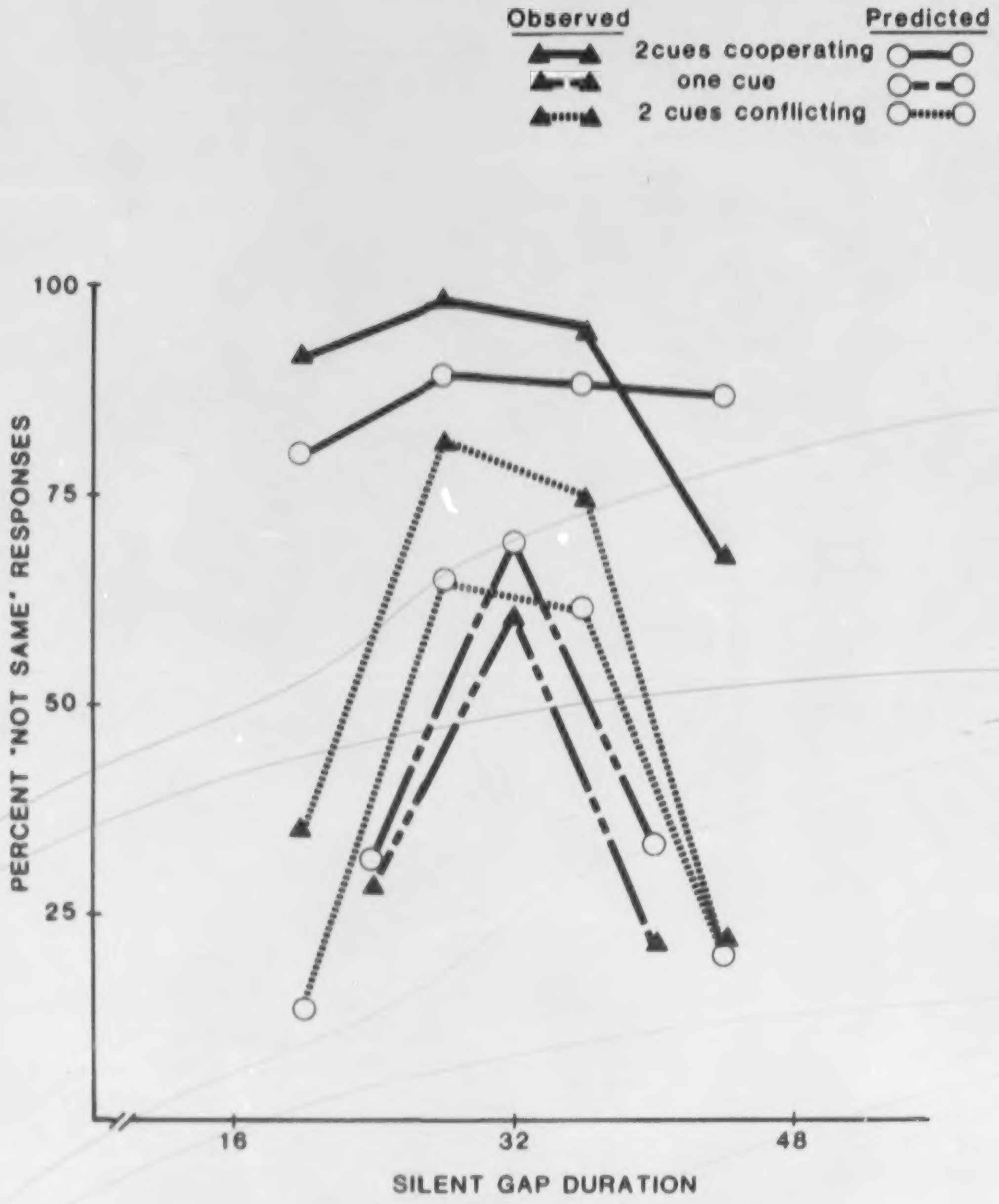


Figure 5. Children's discrimination functions for 2IAX comparisons.

was formed by varying only the spectral cue. This suggests that children and adults differed either in their perception of the temporal cue alone, or in their relative weighting of the temporal and spectral cues for phonetic integration in /s/ + stop cluster perception. The former possibility seems less likely given previous reports that children (e.g., Wolf, 1973) and even infants (e.g., Eimas, Siqueland, Jusczyk, & Vigorito, 1971) show the same VOT boundary (a temporal cue) as adults in perception of stop voicing.

The pattern of results obtained in the discrimination conditions supported the notion that the two acoustic cues are truly equivalent along a single phonetic dimension in children's perception of speech, even though the stimulus pairings used were not ideally suited to the magnitude of the children's trading relation. For the children, both the expected and observed discrimination performances were significantly better when the spectral and temporal cue values "cooperated" phonetically to enhance discrimination along the phonetic dimension, than when the cues "conflicted" phonetically to reduce discriminability along the phonetic dimension. Since the Cooperating and Conflicting Cues conditions involved comparisons that differed by equal amounts along the two acoustic dimensions, the pattern of discrimination findings indicated that the children were not focusing on the acoustic differences as such. Instead, like adults, they perceived the unified phonetic information underlying the diversity in acoustic information.

The cause of the age-related perceptual differences in the magnitude of the trading relation is not directly revealed by this study, and warrants further exploration. One possible reason for the difference might be a lowered sensitivity to frequency differences among formant transition onsets in children vs. adults (Eguchi, 1976); however, the lack of an age effect in the "ay-day" test eliminates the possibility of an absolute age difference in frequency sensitivity for F1 onset values in our stimuli. Children at this age are apparently equal to adults in their perceptual use of a 230- vs. 430-Hz F1 onset difference to signal a difference in degree of alveolar stop closure; that is, they do not differ from adults in their use of that acoustic information as a primary cue to a phonetic distinction. They deviate from adults only in their use of the same acoustic information as a secondary cue to a multiply-cued phonetic contrast. This would suggest that the age difference is more likely related to developmental changes in selective attention to perceptual information than it is to changes in basic auditory sensitivity. It finds converging support from Bernstein's (1979) report that children are less consistent than adults in using F0 as a secondary cue to stop contrasts.

A second possibility is that the age difference in perception of multiple acoustic cues to a phonetic contrast might also relate in some way to child vs. adult production differences. Children six years of age produce shorter VOTs (Kent, 1981), and they show less of a VOT distinction (Kent, 1976) for stop consonants in syllable initial position, relative to adults' productions. Furthermore, children's VOT for stops in /s/ + stop clusters is about 12 msec, averaging across three places of articulation (see Figure 3 in Bond & Wilson, 1980) whereas, in adult production, the average VOT is 23 msec, again averaging across three places of articulation (see Table 1 in Klatt, 1975). Since children produce both word-initial voiceless stops and those following initial /s/, with a shorter VOT than adults, this means that they start phonation earlier after the release of the constriction. In turn, this would imply a lower F1 onset frequency in children's voiceless stops than in adults', at least for those following /s/, and that the F1 onset frequency differences would therefore be smaller for children's voiced-voiceless distinctions in production. The obtained smaller trading relation in the children, for our /s/ + stop cluster, would seem to imply lowered *perceptual* use of the F1 onset distinction, as well as lowered *productive* use of F1 onset distinctions, relative to adults. This hypothesized relation between children's smaller perceptual trading relation and their production of smaller voicing category distinctions could be tested by examining children's gap durations and F1 onsets in "say"-"stay" production relative to their perceptual equivalence tests for "say"-"stay." A relationship between perception and production abilities in 3-year-olds, for example, has been reported for the contrasts /w/, /r/, and /l/ (Strange & Broen, 1981), and has also been indicated by the research of Bailey and Haggard (1980) on voicing distinctions.

Perception of running speech in the natural environment depends upon a listener's ability to integrate multiple acoustic cues, which may interact in complex ways to specify phonetic category information. Yet developmental research on perceptual integration of multiple acoustic cues specifying phonetic content has been sorely lacking. As the results of the present study indicate, examining children's and adults' perception of simple one-cue word-initial differences provides little information about developmental changes in listeners' abilities to integrate and utilize these cues for phonetic perception in multiple-cue contexts, which more closely approximate the diverse information available to a listener in natural speech. In order to better understand developmental changes in the perception of speech it is important that we begin to examine perceptual abilities that more closely approximate those necessary for the perception of speech in the natural environment.

REFERENCES

Abramson, A., & Lisker, L. Voice onset time in stop consonants: Acoustic analysis and synthesis. In D. E. Cummins (Ed.), *Proceedings of the 5th Congress of International Acoustics*. Liege, Belgium: Thone, 1965.

Bailey, P. J., & Haggard, M. P. Perception-production relationships in the voicing contrast for initial stops in 3-year-olds. *Phonetica*, 1980, *37*, 377-396.

Bailey, P. J., & Summerfield, Q. Information in speech: Observations on the perception of [s]+stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, *6*, 536-563.

Bernstein, L. E. Developmental differences in labeling VOT continua with varied fundamental frequency. *Journal of the Acoustical Society of America*, 1979, *65* (Suppl.), S112. (Abstract)

Best, C. T., Morrongiello, B., & Robson, R. Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, 1981, *29*, 191-211.

Bond, Z. S., & Wilson, H. F. /s/ plus stop clusters in children's speech. *Phonetica*, 1980, *37*, 149-158.

Dorman, M. F., Raphael, L. J., & Isenberg, D. Acoustic cues for a fricative-affricate contrast in word-final position. *Journal of Phonetics*, 1980, *8*, 397-405.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, 1979, *65*, 1518-1532

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, 1977, *22*, 109-122.

Eimas, P., Siqueland, E. R., Jusczyk, P., & Vigorito, J. Speech perception in early infancy. *Science*, 1971, *171*, 304-306.

Eguchi, S. Difference limens for the formant frequencies: Normal adult values and their development in children. *Journal of the American Audiological Society*, 1976, *1*, 145-149.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. Perceptual equivalence of two acoustic cues for stop consonant manner. *Perception & Psychophysics*, 1980, *27*, 343-350.

Haggard, M. P., Ambler, S., & Callow, M. Pitch as a voicing cue. *Journal of the Acoustical Society of America*, 1970, *47*, 613-617.

Harris, K. S., Hoffman, H. S., Liberman, A. M., Delattre, P. C., & Cooper, F. S. Effect of third-formant transitions on the perception of the voiced stop consonants. *Journal of the Acoustical Society of America*, 1958, *30*, 122-126.

Kent, R. D. Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research*, 1976, *19*, 421-447.

Kent, R. D. Sensorimotor aspects of speech development. In R. Aslin, J. Alberts, & M. Peterson (Eds.), *Development of perception* (Vol. 1). New York: Academic Press, 1981.

Kewley-Port, D., & Preston, M. Early apical stop production: A voice onset time analysis. *Journal of Phonetics*, 1974, *2*, 195-210.

Klatt, D. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 1975, *18*, 686-705.

Lisker, L. Is it VOT or a first-formant transition detector? *Journal of the Acoustical Society of America*, 1975, *57*, 1547-1551.

Lisker, L., Liberman, A. M., Erickson, D. M., Dechovitz, D., & Mandler, R. On pushing the voice-onset-time (VOT) boundary about. *Language and Speech*, 1977, *20*, 209-216.

Miller, J. L., & Liberman, A. M. Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, 1979, *25*, 457-465.

Repp, B. H. Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech*, 1979, *22*, 173-189.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. Perceptual integration of acoustic cues for stop, fricative and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, *4*, 621-637.

Simon, C., & Fourcin, A. J. Cross-language study of speech-pattern learning. *Journal of the Acoustical Society of America*, 1978, *63*, 925-935.

Strange, W., & Broen, P. The relationship between perception and production of /w/, /r/ and /l/ by three-year-old children. *Journal of Experimental Child Psychology*, 1981, *31*, 81-102.

Summerfield, Q., & Haggard, M. On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. *Journal of the Acoustical Society of America*, 1977, *62*, 436-448.

Wolf, C. The perception of stop consonants by children. Journal of Experimental Child Psychology, 1973, 16, 318-331.

Zlatin, M. A., & Koenigsknecht, R. A. Development of the voicing contrast: Perception of stop consonants. Journal of Speech and Hearing Research, 1975, 18, 541-553.

Zlatin, M., & Koenigsknecht, R. Development of the voicing contrast: A comparison of voice onset time in stop perception and production. Journal of Speech and Hearing Research, 1976, 19, 78-92.

FOOTNOTES

[1] Simon and Fourcin did not test the English-speaking and French-speaking children on the same voicing contrasts, and the contrasts were chosen such that neither group was tested on all three places of stop articulation. The English-speaking children were tested with "coat-goat" (3-14 year-olds) and "Paul-ball" (2-year-olds), whereas the French children were tested with "toto-dodo." Moreover, the children were given only three presentations of each stimulus from a continuum, which is an extremely low number of repetitions (most adult studies use 10-20 presentations per token) and could artificially inflate the children's variance in performance, especially at younger ages.

[2] In American English, the phonetic and articulatory properties of /t/, /p/, or /k/ following /s/ are actually more characteristic of their voiced cognates /d/, /b/, and /g/, respectively. Thus /stei/ with the /s/ noise removed sounds like "day" rather than "tay."

[3] For the two "say"-"stay" continua in Experiment 1 of Best et al. (1981), the /s/ and the synthetic syllable were separated by silent gaps ranging between 0 and 136 msec, in 8 msec increments, resulting in 18 stimuli per continuum. As mentioned in the Introduction, the average trading relation for adult listeners in Experiment 1 of Best et al. (1981) was 24.6 msec. In Experiment 2 of Best et al. (1981) a truncated "say"-"stay" continuum containing 13 stimuli each was used; stimuli containing gaps greater than 96 msec were eliminated, since the adults in Experiment 1 had identified these as "stay" nearly 100% of the time. The average trading relation for adults tested with this truncated "say"-"stay" continuum was 18.5 msec. Because children were tested with the truncated continuum only, our statistics in the present study compared the size of their trading relation relative to the adult trading relation of Experiment 2 (see Figures 2 and 3). However, because the children's discrimination data were obtained prior to completion of testing adults in Experiment 2 of Best et al. (1981), the children's discrimination test was set up based on the adult trading relation of Experiment 1 of Best et al., which was 24.6 msec.

[4] It is interesting, however, that when the "ay-day" data for individual children were compared to the magnitudes of their "say-stay" trading relations, there was a tendency for children with larger-magnitude trading relation to also show larger differences in percent "day" identifications between the "weak day" and "strong day" syllables.

[5] Although the order of the observed peaks across the three discrimination conditions matched the order of the predicted peaks, there was some discrepancy between observed and predicted levels of performance. There was no performance difference between observed and predicted scores across the One Cue comparisons, but there was a significant main effect for observed vs. predicted across the Conflicting Cues comparisons, $F_{1,17} = 18.7$, $p < .005$, and across the Cooperating Cues comparisons, $F_{3,21} = 5.8$, $p < .005$. T-tests comparing observed and predicted performance obtained performance to be marginally better than predicted for all Conflicting Cues comparisons, and for the Cooperating Cues comparisons that involved stimuli from the "stay" identification category. These moderate differences in obtained vs. predicted performance levels indicate some ability to discriminate acoustic differences between stimuli beyond differences in phonemic identity. However, this is not particularly damaging to the phonetic perceptual equivalence hypothesis since the observed-predicted differences are similar in magnitude to those found in adults by Best et al. (1981), and in fact are common in studies on categorical perception of speech segment contrasts.

THE ROLE OF THE STRAP MUSCLES IN PITCH LOWERING

Donna Erickson, Thomas Baer, and Katherine S. Harris+

INTRODUCTION

It has long been recognized that the extrinsic laryngeal muscles may participate in the control of fundamental frequency (F_0) during singing or speech. There is a large body of direct physiological evidence for this participation for the case of singing (e.g., Faaborg-Anderson & Sonninen, 1960). However, there are several reasons to expect that the extrinsic muscles are also involved in F_0 control--especially for F_0 lowering--during speech production. Recent studies of laryngeal control of F_0 falls in speech have implicated the cricothyroid and the strap muscles as the primary muscles involved in F_0 lowering (e.g., Atkinson, 1978; Erickson, 1976; Erickson & Atkinson, 1976; Simada & Hirose, 1971). Specifically, the cricothyroid shows decreased activity and the strap muscles increased activity during pitch falls. In this paper, we wish to examine the interaction between the cricothyroid and strap muscles in effecting F_0 fall in more detail, and in particular, to study their joint activity.

BACKGROUND

During speech or singing, fundamental frequency is determined primarily by activity of the intrinsic laryngeal muscles, and, to a lesser extent, by subglottal pressure (Baer, 1979; Hixon, Klatt, & Mead, 1971). Given that the vocal folds are in a voicing position (partially or fully adducted), and that sufficient subglottal pressure to maintain phonation has been produced, F_0 is determined to a substantial degree by the tension of the vocal folds, which is, in turn, determined by adjustments of the relative positions of the cricoid, thyroid, and arytenoid cartilages. Recent results have unanimously shown that the muscle whose activity is most directly related to F_0 is the cricothyroid (CT), a finding consistent with the anatomical fact that the cricothyroid muscle is best suited for increasing the distance between the anterior part of the thyroid cartilage and the arytenoid cartilages. The only muscles that could shorten this distance by action at the level of the folds themselves, however, are the laryngeal sphincter muscles--the thyroarytenoid (TA), and the muscles of the aryepiglottic sphincter. Of these, it is known that the activity of the internal part of the thyroarytenoid (the vocalis) is not usually positively correlated with F_0 lowering (Gay, Hirose, Strome, & Sawashima, 1972; Shipp & McGlone, 1971). Thus, if an active shortening-

---

+Also The Graduate School, City University of New York.

Acknowledgment. This work was supported by NINCDS Grants NS-13617 and NS-13870, and BRS Grant RR-05596 to Haskins Laboratories.

lowering mechanism exists, it must either involve the external part of the TA muscle or some more indirect action through the aryepiglottic sphincter muscles, or the action of the extrinsic laryngeal muscles.

Untrained singers allow the whole larynx to move upward during increases of F_0 and downward for decreases of F_0. There is also some evidence that similar tendencies occur, on the average, during speech intonation (Ewan, 1979). Since the vertical position of the larynx as a whole is determined by its extrinsic attachments, this constitutes evidence that the extrinsic muscles are activated with changes in F_0. There is direct electromyographic and clinical evidence that the extrinsic muscles are involved in the production of both the high and low extremes of a singer's F_0 range (Sonninen, 1956). Since the range of fundamental frequency employed during speech production usually lies near the low extreme of singing range, we might expect the extrinsic muscles to participate in F_0 lowering during speech.

A knowledge of the anatomy of the region and those few experimental facts available have been used to develop a number of theories to account for F_0 lowering; among these are (1) the passive relaxation theory (Zemlin, 1959), (2) the external frame function theory (Sonninen, 1956), (3) the vertical tension theory (Ohala, 1972), and (4) the laryngeal articulation theory (Lindqvist, 1972). In the first, the passive theory, F_0 lowering is said to result simply from relaxation of the F_0 raising musculature (i.e., cricothyroid) with no active gesture. In the second, the external frame function theory (which is the one we will be most concerned with here), F_0 lowering is thought to be brought about by a horizontal shortening of the vocal folds due to forces exerted by the external attachments to the larynx. In the third, the vertical tension theory, F_0 lowering is said to result from a lowering of the larynx; that is, the vertical height of the larynx is related to F_0 directly through vertical stretching of the surface membranes of the larynx, rather than by horizontal lengthening as in the external frame function theory. In the fourth, the laryngeal articulation theory, F_0 lowering is said to be brought about by the laryngeal and supra-laryngeal sphincter muscles opposing the cricothyroid muscle, so that both vocal fold shortening and supraglottal constriction result.

It is possible that several of the theories listed above may be "correct." That is, each of the possible mechanisms might be used at different times or in different combinations. However, it is clear that there are changes in the activity of the extrinsic muscles during speech production and that these muscles are capable of changing the configuration of the laryngeal cartilages.

Figure 1 shows a schematic side view of the larynx, indicating the major structures and their attachments. The three major structures important for F_0 control are the cricoid cartilage, the thyroid cartilage, and the hyoid bone. Because of the ligamentary and muscular attachments between these three structures, movement of any one of them produces changes in the forces exerted on the other two, in general causing them to move also. Each of the three structures also has attachments to other body structures. Therefore, any movement causes a readjustment of the forces not only that the three structures exert on each other, but also that external attachments exert.

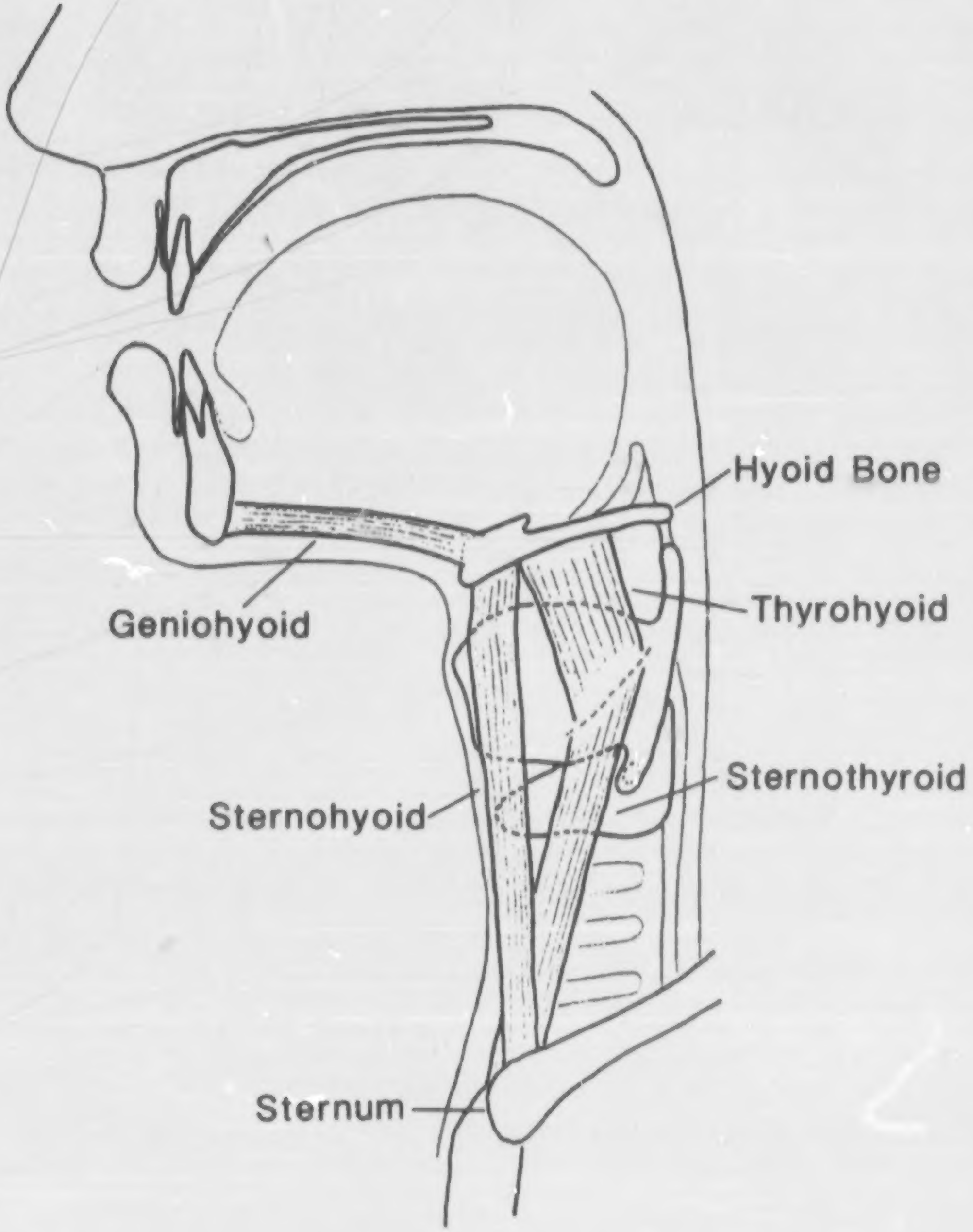

Hyoid Bone
Geniohyoid
Thyrohyoid
Sternothyroid
Sternohyoid
Sternum

Figure 1. Lateral view of larynx and supra-laryngeal structures.

Specific theories of strap muscle action must be assessed within the framework of this biomechanical complexity. For example, Sonninen (1956), who simulated muscle action by pulling individual muscles in cadavers fixed in various head positions, found that a pull on the sternothyroid (ST) caused the thyroid cartilage to move and tilt forward slightly. Due to the attachment of cricoid and thyroid cartilage, a tug on one caused a movement of the other. "Contraction" of ST resulted in either lengthening or shortening of the vocal folds depending on the position of the head and cervical spine: "If the tilting of the cricoid cartilage exceeded that of the thyroid cartilage, the vocal cords shortened, if it was less, they lengthened" (p. 25).

While Sonninen believed, on anatomical grounds, that this anterior vector of movement might result from the contraction of any of the three strap muscles, i.e. the sternohyoid, the sternothyroid and the thyrohyoid, whether or not a vertical component was present, he did not investigate the problem. Later investigators have been in disagreement as to whether there is functional differentiation among the muscles. Collier (1975) and Hiki and Kakita (1976) report a difference, although Erickson (1976) does not. Moreover, the last-named study shows that all three straps appear to be associated with F_0 lowering in the low part of the F_0 range.

In the articles cited above, investigators have not always differentiated what is biomechanically possible from what is actually used as a maneuver for pitch control by speakers or singers. Further, speakers may differ from trained singers in what they do. In the study that follows, we have tried to look at reasonably common mechanisms in speakers without special training whose language calls for precise control. Hence, we have used speakers of Thai, a tone language, as subjects and compared them with speakers of English.

DESCRIPTION OF EXPERIMENTAL STUDY

In order to assess the role of strap muscles in F_0 lowering, we performed the following experiment with two Thai and two English speakers on utterances that showed falling F_0 contours.

We used the EMG and F_0 processing facilities at Haskins Laboratories and restricted our study to the cricothyroid (CT) muscle and the strap muscles. As mentioned earlier, there is no strong evidence for a differentiation among the strap muscles. But since the earlier literature, especially Hirano, Ohala, and Vennard (1969), has focussed attention on the sternohyoid (SH), we have given it special attention. However, in the case of Thai speaker PT, since SH proved not to be a good insertion, we examined the thyrohyoid (TH) muscle. The muscle insertions were performed by Hajime Hirose, using insertion techniques he has described (Hirose, 1971).

In Thai, we examined F_0 falls on words with two types of tones, the "falling" tone and the "low" tone, i.e., /bâa/, /bîi/, /bûu/, and /bàa/, /bìi/, /bùu/. The words were spoken in a carrier phrase /əə-/, meaning "Yes, that is a _____." In Thai, these two tones begin their fall at a relatively high value of F_0, or a mid value, respectively.

In English, we examined falling F_0 contours from the words "Bev" and "loves" in the sentence "Bev loves Bob," with emphatic stress on one of the three words. The word is produced with intonation that falls from a high value, if it is stressed, or a mid value, if it is not. The particular samples used are those described in Atkinson (1973, 1978). We will describe the two types of F_0 falls in the two languages as "high falls" and "mid falls."

The two speakers for the English sentences were one native American (male), and one naturalized American (male) whose native language was Estonian, but who was a fluent[1] speaker of English. The speakers for the Thai sentences were two native speakers (male) of the central dialect of Thailand (as spoken in Bangkok) who were students at the University of Connecticut. The two English speakers were sophisticated with respect to the literature on F_0 control: the two Thai speakers were not.

Previous studies (e.g., Atkinson, 1973; Erickson, 1976) indicate a typical pattern of CT and SH activity occurring with falling F_0 contours. Prior to the fall in F_0 the CT shows a decrease in activity, and after the fall, the SH shows an increase in activity. In order to determine whether the CT and SH could be in some way causing the fall in F_0, we examined the delay between onset of F_0 fall and onset of the decrease in CT activity on the one hand, and onset of the increase in SH activity on the other hand. This method was first reported in Atkinson and Erickson (1977) and Erickson (1976).

Schematic patterns of F_0, CT, and SH strap muscle activity are shown in Figure 2. The onset of F_0 fall is fairly abrupt, and easily determined by visual inspection for measurement purposes. The onset of strap muscle activity was also fairly easy to determine, since usually there was a low steady base level of activity, followed by a sudden increase. It was at the point where the EMG curve began to increase that the measurements were made for the strap muscles. The cricothyroid showed a clear peak or peaks of activity before it sloped off into a steady low level pattern of activity. It was at the point where the EMG curve began to descend that the measurements were made. We examined individual tokens of each of the four speakers: 30 tokens each for the Thai speakers, and 20 tokens each for the English speakers. Tokens in which clear peaks were not observed were discarded.

RESULTS

The distribution of delay times between the change in EMG activity and F_0 fall for high falls is shown in Figure 3. All four speakers show a pattern in which CT activity generally begins to decrease before F_0 fall. For three of the four speakers, strap muscle activity follows the onset of F_0 fall, while for the fourth, KO, it precedes it.

The data for the first three speakers suggest the following: (1) Since the CT is active prior to the F_0 fall, it is certainly possible that relaxation of the CT could be causal with regard to the initiation of F_0 lowering (2) Since the strap muscle is not active until after the F_0 fall, it is clearly not possible for the strap muscles to be causal with regard to the initiation of F_0 lowering.

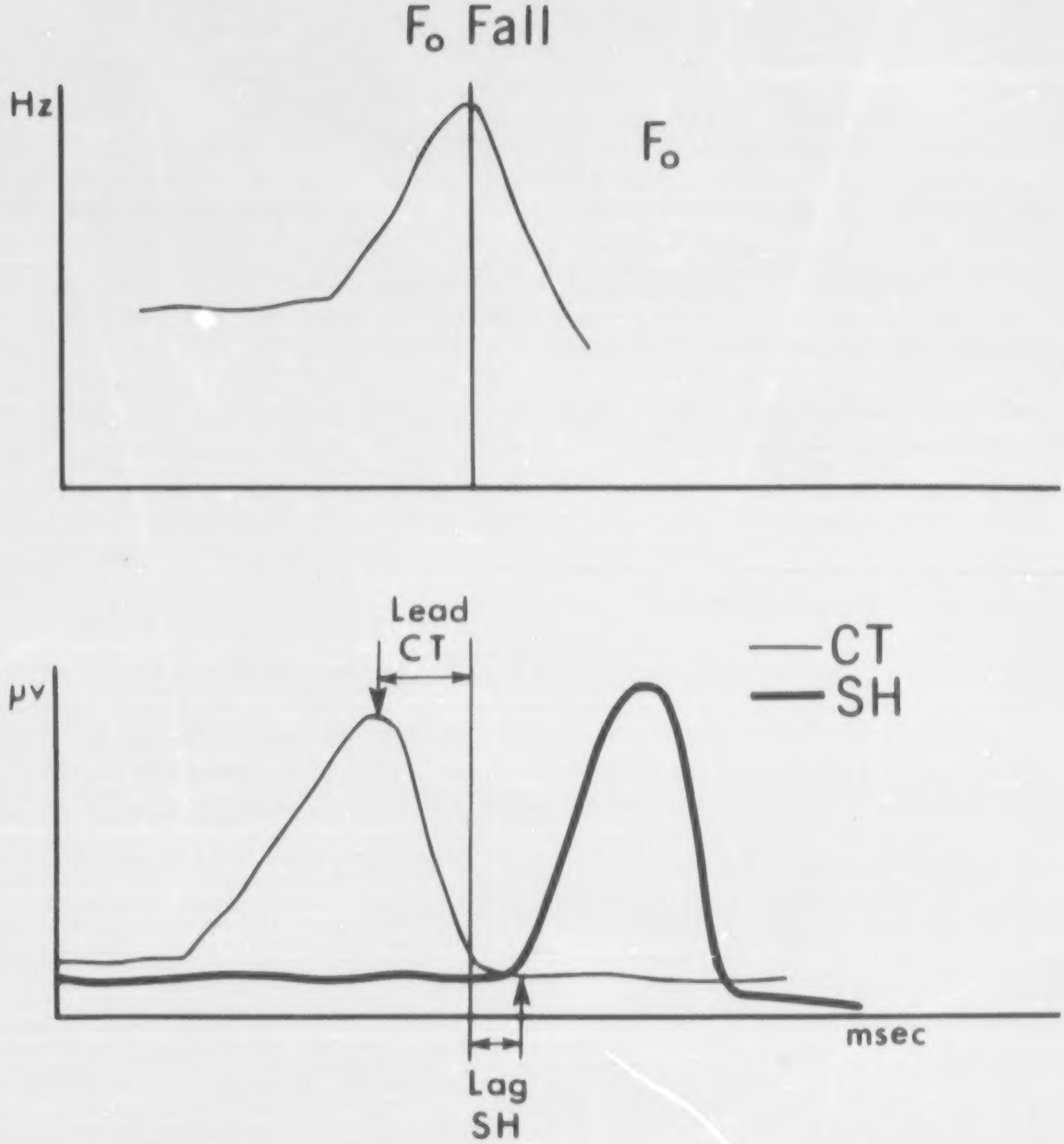

Fo Fall
Hz
Fo
Lead
CT
μv
CT
SH
msec
Lag
SH

Figure 2. Schematic representation of cricothyroid and strap muscle activity in relation to F_0 fall.

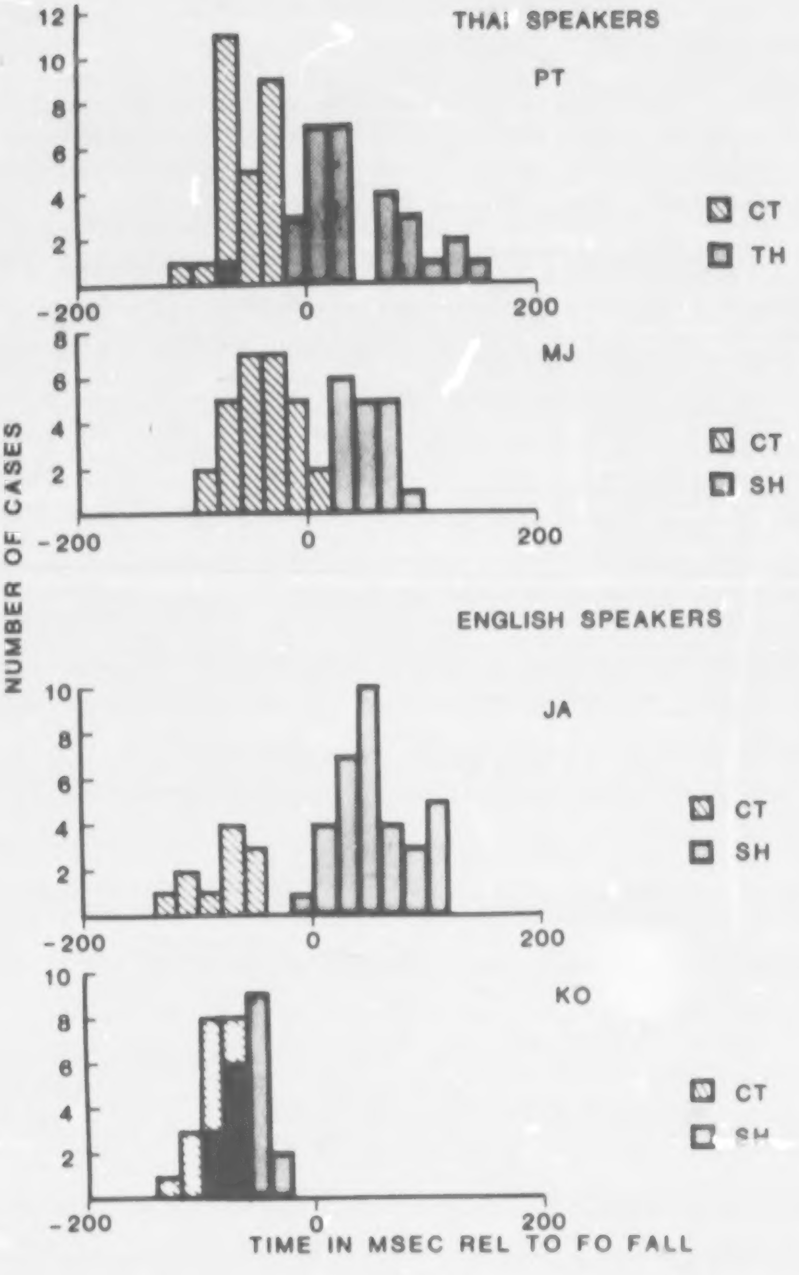


Figure 3. Data for high falls. Change in activity for cricothyroid and strap muscle activity in relation to F_0 fall.

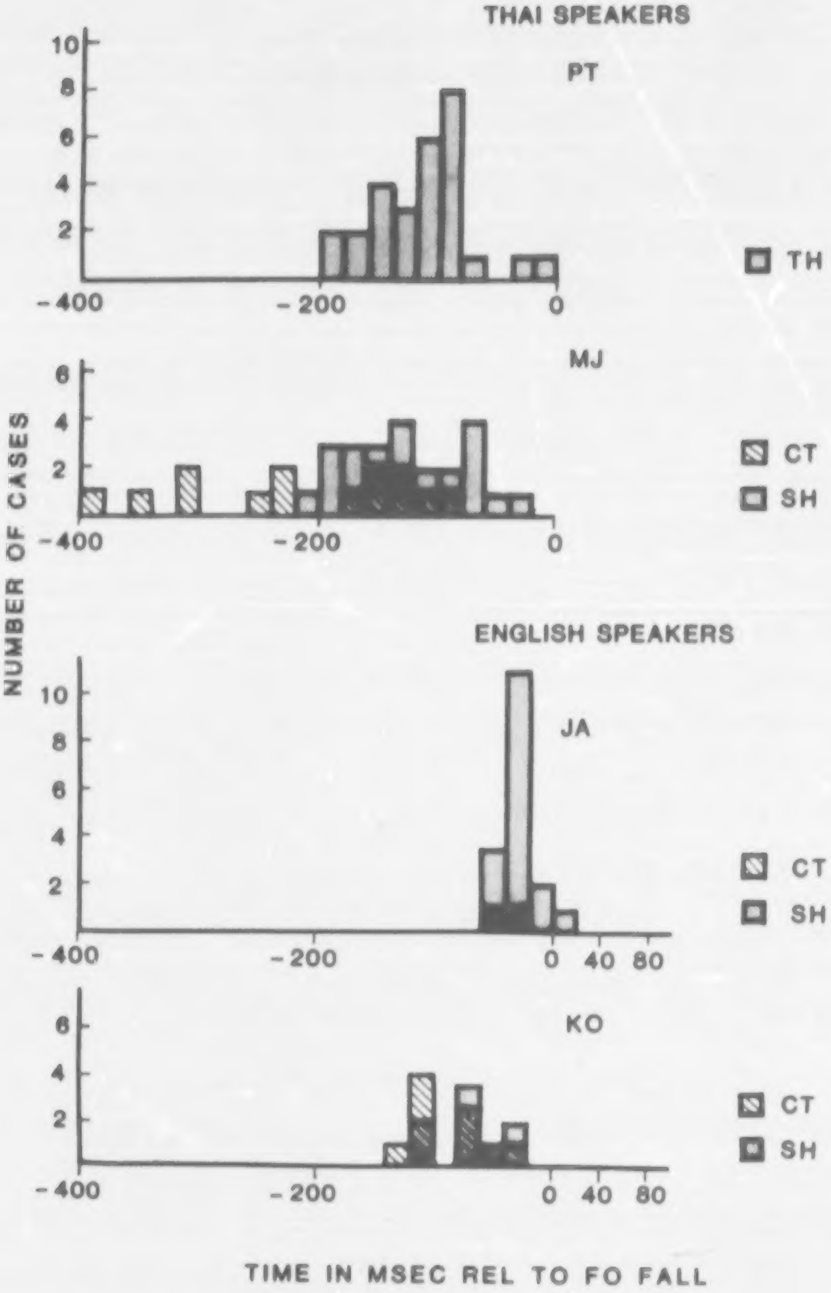


Figure 4. Data for mid falls. Change in activity for cricothyroid and strap muscle activity in relation to F_0 fall.

For the fourth speaker, KO, who does not follow the above pattern, we can conjecture that (1) the CT is probably causal with regard to the initiation of F_0 lowering, but (2) whether the strap muscle is causal also is not at all clear. The data on speaker KO may reflect an alternative F_0 lowering strategy.

Next, we consider patterns of F_0, CT, and strap muscle activity for mid-fall situations shown in Figure 4. In comparison with the patterns for high fall situations previously described, we note initially, that the cricothyroid muscle tends to show less dynamic changes in activity in mid falls than in high falls. This pattern has also been found by other experimenters. For instance, Rubin (1963) noted that CT activity is "virtually absent in lower frequencies, minimum just above this, and does not really become intense until transition to the middle register" (p. 1002). Given that the transitions in CT activity are far less abrupt for mid- to low-falls, we were not able to establish onset or offset points as readily. Hence, the number of cases for the mid-fall distributions is much smaller.

In examining the delay time measurements for mid-falls, displayed in Figure 4, we see the following pattern of strap muscle activity: Strap muscle activity starts to increase before the initiation of the F_0 fall. This contrasts strongly with the pattern of strap activity seen with high falls, where strap activity begins after initiation of F_0 fall.

The findings reported in this study lend themselves to certain interpretations concerning how the laryngeal muscles work to lower F_0. For one thing, it is obvious that CT and strap muscles act synergistically in lowering F_0. Simply speaking, the CT must be relaxed (or relaxing) before the strap muscles can participate in F_0 lowering. A more complicated statement emerges when we compare the patterns of CT-strap muscle activity for the two types of fall situations, i.e., high to low, and mid to low falls. A fall from high to low F_0 is initiated by relaxation of the CT, with the strap muscles showing activity well after initiation of the F_0 fall. However, a fall from mid to low F_0 is initiated by the strap muscles, with the CT playing relatively little role.

REFERENCES

Atkinson, J. E. Aspects of intonation in speech: Implications from an experimental study of fundamental frequency. Unpublished doctoral dissertation, University of Connecticut, 1973.

Atkinson, J. E. Correlation analysis of the physiological features controlling fundamental voice frequency. Journal of the Acoustical Society of America, 1978, 63, 211-222.

Atkinson, J. E., & Erickson, D. The function of strap muscles in speech: Pitch lowering or jaw opening? Haskins Laboratories Status Report on Speech Research, 1977, SR-49, 97-102.

Baer, T. Reflex activation of laryngeal muscles by sudden induced subglottal pressure changes. Journal of the Acoustical Society of America, 1979, 65, 1271-1275.

Collier, R. Physiological correlates of intonation patterns. Journal of the Acoustical Society of America, 1975, 58, 249-255.

Erickson, D. M. *A physiological analysis of the tones of Thai*. Unpublished doctoral dissertation, University of Connecticut, 1976.

Erickson, D., & Atkinson, J. E. The functions of the strap muscles in speech. *Haskins Laboratories Status Report on Speech Research*, 1976, *SR-45/46*, 205-210.

Ewan, W. G. Laryngeal behavior in speech. *Report of the Phonology Laboratory No. 3*. University of California at Berkeley, 1979.

Faaborg-Anderson, K., & Sonninen, A. The function of the extrinsic laryngeal muscles at different pitch. *Acta Oto-laryngologica*, 1960, *51*, 89-93.

Fujimura, O. Control of the larynx in speech. *Phonetica*, 1977, *34*, 280-288.

Gay, T., Hirose, H., Strome, M., & Sawashima, M. Electromyography of the intrinsic laryngeal muscles during phonation. *Annals of Otology, Rhinology and Laryngology*, 1972, *81*, 401-409.

Hiki, S., & Kakita, S. A study on laryngeal control for pitch change by use of anatomical structure model. *Conference Record, IEEE International Congress on Acoustics, Speech, and Signal Processing*, 1976, 43-46.

Hirano, M., Ohala, J., & Vennard, W. The function of the laryngeal muscles in regulating fundamental frequency and intensity of phonation. *Journal of Speech and Hearing Research*, 1969, *12*, 616-628.

Hirose, H. Electromyography of the articulatory muscles: Current instrumentation and techniques. *Haskins Laboratories Status Report on Speech Research*, 1971, *SR-25/26*, 73-86.

Hixon, T. J., Klatt, D. H., & Mead, J. Influence of forced transglottal pressure changes on vocal fundamental frequency. *Journal of the Acoustical Society of America*, 1971, *49*, 105.

Lindqvist, J. Laryngeal articulation studied on Swedish subjects. *Quarterly Progress and Status Report of the Speech Transmission Laboratory* (Royal Institute of Technology, Stockholm), 1972, *2/3*, 10-27.

Ohala, J. How is pitch lowered? *Journal of the Acoustical Society of America*, 1972, *52*, 124. (Abstract)

Rubin, H. J. Experimental studies of vocal pitch and intensity in phonation. *Laryngoscope*, 1963, *73*, 973-1025.

Shipp, T., & McGlone, R. E. Laryngeal dynamics associated with voice frequency change. *Journal of Speech and Hearing Research*, 1971, *14*, 761-768.

Simada, Z., & Hirose, H. An EMG study of Japanese accent patterns. *Proceedings of the 7th International Congress of Acoustics*, 1971, *13*, 121-124.

Sonninen, A. The role of the external laryngeal muscles in length adjustment of the vocal cords in singing. *Acta Oto-laryngologica*, 1956, *Suppl. 130*.

Zemlin, W. R. *A comparison of a high speed cinematographic and a transillumination photo conductive technique in the study of the glottis during voice production*. Unpublished Master's thesis, University of Minnesota, 1959.

FOOTNOTE

[1] The subject's speech was marked by some foreign interference. While he was not an ideal subject, he was the only volunteer for what seemed at the time (1973) a fairly formidable procedure. However, his productions were perceptually normal as to intonation contour and the interest here is not in the choice of English, but of any non-tone language.

PHONETIC VALIDATION OF DISTINCTIVE FEATURES: A TEST CASE IN FRENCH[*]

Leigh Lisker[+] and Arthur S. Abramson[++]

Abstract. Much of the phonological literature shows little concern for recent phonetic data. Even in a provocative overview of Jakobsonian phonology (Jakobson & Waugh, 1979) that does give much attention to recent phonetic research, the latter is not exploited very convincingly in defining certain distinctive features. A case in point is the notorious French chestnut embodied in *vous la jetez* vs. *vous l'achetez*, a pair of expressions traditionally said to be distinguished by a voicing feature in the palatal fricatives, which appear here as initial elements in consonant clusters with /t/. It is reported, however, that the /ʒ/ of *jetez* is devoiced through assimilation to the following /t/, and it is argued that a feature of "fortisness" or "tensity" is therefore needed. We have tested two hypotheses: (1) Such pairs are likely to be distinguished in production and perception; (2) When they are distinguished, the phonetic basis is glottal adduction vs. abduction. Readings by native speakers of standard French of written sentences terminating in *la jeter* and *l'acheter* were collected and those tokens in which the terminal items were pronounced as disyllables were presented to French listeners for identification. Their responses suggest instability of the distinction, with a perceptual bias toward /ʃ/, thus largely negating the first hypothesis. Insofar as the distinction *is* maintained, spectrographic analysis and perceptual tests involving the manipulation of /ʒ/ and /ʃ/ noise segments do not argue against a hypothesis of laryngeal control.

If phonology is to be taken seriously as more than an elaborate spelling exercise--in other words, if the assertions of phonetic fact are not just objects to be manipulated rather than statements whose truth values are thought relevant to linguistic description, then they deserve the respect implied by careful and appropriate testing. Terms such as "voiced" and "fricative" have physical meanings that are generally recognized. Provided that the linguist who says that a given utterance type involves a voiced fricative grants physical meaning to those terms, the statement may be checked against physical observation. Linguists may not want to test their phonetic judgments, even though ostensibly they are making claims about the physical

---

[*] Paper presented at the 56th Annual Meeting of the Linguistic Society of America, New York City, December 1981.

[+] Also University of Pennsylvania.

[++] Also University of Connecticut.

Acknowledgment. This work was supported by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-70 (1982)]

nature of speech signals. Quite frankly we find such an attitude deplorable, even if we acknowledge that beliefs about the nature of the world are also facts worth studying. Some kinds of phonetic judgments are, moreover, not easily translated into terms that allow ready testing. An outstanding example is the claim that two utterance types are distinguished by a difference in force of articulation, where the so-called "fortis-lenis" distinction is attributed to particular segments. It might be argued that if the fortisness of a particular segment is a matter of belief that is widely shared, then it may not be dismissed as groundless just because laboratory phoneticians have failed to find an appropriate measure. But there is a difference between taking such a belief seriously and regarding it as sacrosanct. We prefer to take it seriously, and that means to view it critically.

The claim that a phonological distinction is based on a fortis-lenis difference is not easily tested for another reason, namely because most often a non-controversial difference is present, one that *is* physically interpretable. Only rarely is an alleged fortis-lenis difference unaccompanied. One of these cases seems to be in French, a language that distinguishes two sets of obstruents, one usually voiced and the other voiceless. A number of linguists (e.g., Armstrong, 1932; Delattre, 1941; Malmberg, 1943), most recently Jakobson and Waugh (1979) have said that the palatal fricatives /ʒ/ and /ʃ/, usually voiced and voiceless respectively, are lenis and fortis as well. They claim, moroever, that in the phrase *Vous la jetez* 'You throw it' a common pronunciation omits the schwa that in a more deliberative style separates the /ʒ/ and the /t/, and also devoices the fricative. The resulting form, it is further said, is distinguishable from the semantically different expression *Vous l'achetez* 'You buy it,' despite the alleged absence of any voicing difference. The aim of the exercises to be reported here was to test the proposition that the distinction just described cannot be attributed to a difference in laryngeal action, and that we must look for something else that can plausibly be regarded as a consequence of a difference in articulatory force. The strongest acoustic evidence for a difference in laryngeal management would be the presence of glottal pulses during the fricative noise of /ʒ/, and the absence of same during the /ʃ/ noise. The acoustic indices of articulatory force that are commonly proposed are duration and intensity level, in this case the relative durations and intensities of the /ʒ/ and /ʃ/ noises. (It must be pointed out that, on the one hand, the absence of glottal pulses during the /ʒ/ noise does not conclusively demonstrate that the laryngeal action is the same for /ʒ/ and /ʃ/, while a difference in either noise duration or intensity may as plausibly be attributed to a difference in laryngeal management as to one of articulatory force.)

Three tests were run: first, native speakers of French recorded a set of sentences read from a written list, and the recordings were played back to French listeners for identification of the intended target forms; second, selected sentence tokens were edited so that fricative intervals from well-identified *jeter* and *acheter* were interchanged; finally, the intensities of the fricative intervals were varied to determine whether this would affect listeners' identifications of the sentences.

The first test was run just to make sure that sentences meant to differ only as to whether they contained *jeter* or *acheter* could be distinguished if pronounced with fricative-stop clusters. Three speakers of standard French were recorded in readings of the following sentences. The sentences were listed in a random order.

Il faut la jeter.
Il faut l'acheter.
Il ne faut pas la jeter.
Il ne faut pas l'acheter.
Il devrait la jeter.
Il devrait l'acheter.
On a fini par la jeter.
On a fini par l'acheter.
Elle a fini par la jeter.
Elle a fini par l'acheter.
J'ai décidé de la jeter.
J'ai décidé de l'acheter.
Elle ne pouvait pas la jeter.
Elle ne pouvait pas l'acheter.
Est-ce que vous voulez la jeter?
Est-ce que vous voulez l'acheter?
On dit que vous voulez la jeter.
On dit que vous voulez l'acheter.
Moi, j'ai peur de la jeter.
Moi, j'ai peur de l'acheter.
Moi, je ne veux pas la jeter.
Moi, je ne veux pas l'acheter.
Est-ce que vous ne voulez pas la jeter?
Est-ce que vous ne voulez pas l'acheter?

One speaker read all the sentences containing jeter with this word pronounced as a disyllable. Since her productions could not be used to test our hypothesis, they were discarded. A second speaker always pronounced jeter as a monosyllable, while the third nearly always did so. Randomizations of the sentences recorded by these latter two speakers were played back to native listeners, both the speakers and others. The listeners' judgments as to the identity of the final words (if you like, their judgments as to the speakers' intentions) are presented in Table 1. Speaker G.P., who pronounced all his tokens of jeter as monosyllables, very clearly produced sentences that were ambiguous; roughly two thirds of both intended jeter and acheter were judged to be the latter by the three listeners who rendered a total of 280 responses. In the case of D.E.'s readings, although intended acheter were more often reported as acheter than were intended jeter, it can hardly be said that the 704 responses by four listeners provide strong evidence that the /ʒ/-/ʃ/ distinction can survive deletion of the schwa of jeter. D.E.'s intended jeter were so identified just at chance; her acheter tokens, reported 60% as acheter, were perhaps more often produced with fully voiceless fricative-stop clusters, combinations that might predispose listeners to report acheter. Chi-square tests of the individual listener's responses revealed only a single case in which a speaker's intended forms were correctly identified at better than chance: D.E. as listener was able to identify her own recorded sentences at a level better than p < .001.

The data of our first test suggest that there is little basis, at least for these speakers and listeners, for the claim made as to the robustness of

---

Table 1

Labeling of Original Recordings

Speaker: G.P.

3 listeners

280 responses

	Reported	
Intended	jeter	acheter
jeter	34%	66%
acheter	30%	70%

Speaker D.E.

4 listeners

704 responses

	Reported	
Intended	jeter	acheter
jeter	51%	49%
acheter	40%	60%

---

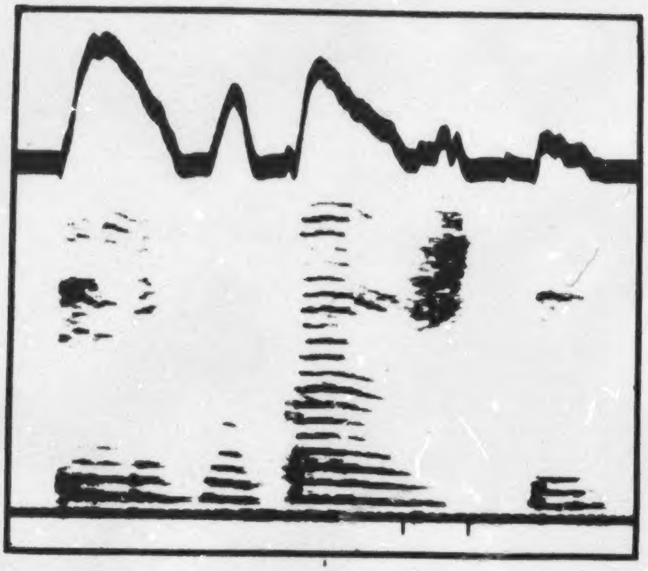
the /ʒ/-/ʃ/ contrast in the context under study. The fortis-lenis difference, so hard for the laboratory phonetician to lay hands on, seems to be no less elusive for our French speakers and listeners. Of course, while our test subjects are certifiably native speakers of French, and the claim is about French, somewhere there may be whole communities of speakers who behave as the claim we are testing says speakers of French do generally. But at the moment we do not know whether or where they are to be found.

At this point we might have dropped the whole matter. We were persuaded to continue, however, by the following consideration. If we could find *any* sentence tokens with intended *jeter* that were so identified, and that we could say were produced in accord with the schwa-deletion rule, and if we also found other tokens regularly judged to contain *acheter*, then we might still pose the original question: does a difference in labeling responses require us to recognize a phonetic basis other than laryngeal? Of the more than 40 sentences that D.E. recorded containing intended *jeter*, just three were reported, at 90% or better, as ending with *jeter*. Of an equal number of tokens with intended *acheter* there were six that were as often so reported.

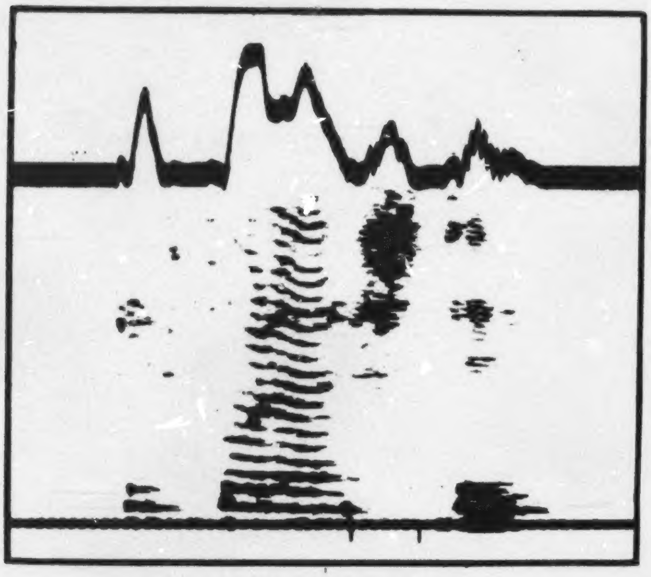
Our data do not compel the conclusion that these particular tokens reflect real auditory/phonetic differences, since purely random labeling behavior might have yielded the results obtained. On the other hand, we cannot absolutely reject the possibility that these *jeter* and *acheter* tokens differ acoustically in a way that can explain why listeners reported them differently. We proceeded therefore to examine spectrographically all the unambiguously labeled sentence tokens, looking for differences that might consistently distinguish members of the two sets, and, if such were to be found, determining whether they were of laryngeal or extra-laryngeal origin.

Figure 1 reproduces narrow-band spectrograms of two sentence tokens with well-identified *jeter* and *acheter*. The short vertical lines at the base of each spectrogram mark off the fricative noise intervals. The two intervals differ very little in duration (perhaps 5%), but they do differ in two other aspects. The amplitude profile for the fricative of *acheter* has a higher peak value, and this is as proponents of a fortis-lenis distinction would predict, although it is also consistent with the higher airflow that should result from the abduction of the vocal folds that occurs in voiceless fricatives. The other difference is in the extent to which the harmonic pattern that characterizes both signals just before the fricative intervals persists past the onset of the noise. In the upper spectrogram of Figure 1 the harmonics fill well over half the fricative interval; in the lower one they damp out much earlier. The spectrograms do not tell us whether amplitude or voicing is perceptually significant, but they suggest that perhaps one or both of them may play some role.

In order to see whether the category assignments of the items differently labeled can be ascribed to the fricative segments, we selected four sentence tokens, two for each reported word, for further testing. For each token the fricative segment was first excised with the help of a waveform editing program, and then each of the four segments was in turn introduced into the gaps left in each of the sentences. The 16 acoustically different signals were then presented in random order to three of our French listeners. Their responses are represented in Table 2. Each number in the table represents the



Il ne faut pas la jeter.



Il faut l'acheter.

Figure 1. Narrow-band spectrograms of sentences with well-identified tokens of jeter and acheter. The short vertical lines mark off the fricative noise intervals.

Table 2

Responses to Cross-Matched Fricative Noises

Speaker: D.E.

3 listeners

192 responses

Noise From

Intended	jeter		acheter
		jeter	
jeter	77%		35%
		acheter	
acheter	50%		75%

Table 3

Responses to Fricative Noises at Two Intensity Levels

Speaker: D.E.

3 listeners

160 responses

		jeter	
Intended	0dB		+10dB
jeter	75%		80%
		acheter	
	-10dB		0dB
acheter	73%		85%

---

averaged responses to four stimuli. For example, the four combinations of the two /ʒ/ noise segments and the two contexts that originally included those segments elicit an average of 77% *jeter* identifications. The four combinations of those same contextual signals with /ʃ/ noises elicited, on the average, only 35% *jeter* judgments. Combinations of /ʃ/ noises with their proper contexts were reported 75% as containing *acheter*. The same contexts with /ʒ/ noise yielded stimuli that were quite ambiguous.

When the responses of each listener were submitted to a simple Chi-square test of significance, only one was found to distinguish reliably between the two classes of stimuli ($p < .001$). Possibly it is significant that this listener was the speaker D.E. The fact that two of our three listeners failed to distinguish two categories makes still more doubtful the proposition that *jeter* and *acheter* maintain phonetic distinctiveness in contexts of the kind tested, in the absence of the schwa that elsewhere marks *jeter*, *even if there seem to be differences in the extent to which voicing accompanies frication*. The fact that the percentage "correct" scores obtained were lower than the 90% obtained for the test tokens in the initial labeling test is not readily explained, but it can be pointed out that three of the four stimuli on which each of the values given in Table 2 is based were "unnatural" combinations of frication noises and sentence contexts, and the process of cutting and recombining may well have introduced incongruities of intensity, duration and fundamental frequency that could contribute to listener uncertainty.

Our last test involved no commutation of segments. Instead, the four noise intervals were presented in their native contexts, but at two intensity levels. In the *acheter* sentences the fricative segments were played back at their original levels and also with 10 dB attenuation. The corresponding segments in the *jeter* sentences were also replayed at their original intensities, and at intensities 10 dB higher. As Table 3 shows, the effects of modifying the intensities of these segments are not spectacular; *acheter* responses decreased little more than 10% with decreased noise intensity, while *jeter* responses actually increased with increased intensity possibly reflecting the effect of the increased salience of the voicing harmonics. Chi-square tests of the responses of the four listeners who underwent this test showed that varying the noise intensities had no statistically significant effect on labeling behavior.

To conclude, we have little reason, on the basis of the data gathered in the course of this study, to believe that speakers of standard French reliably maintain the contrast between a sentence pair *vous la jetez* and *vous l'achetez* in the absence of differences of vocalization and voicing. Thus the alleged basis for an independent fortis-lenis contrast in French seems to us to be very possibly entirely illusory. However, even *if* sporadically we find well-identified fricative-stop clusters that hint at a contrast, we find no compelling evidence to reject an explanation in terms of a difference in laryngeal behavior.

REFERENCES

Armstrong, L. E. *The phonetics of French*. London: Bell and Sons, 1932.

Delattre, P. C. La force d'articulation consonantique en français. *The French Review*, 1941, *14*, 220-232.

Jakobson, R., & Waugh, L. *The sound shape of language*. Bloomington: Indiana University Press, 1979.

Malmberg, B. *Le système consonantique du français moderne*. Lund: Gleerup, 1943.

ON CONSONANTS AND SYLLABLE BOUNDARIES*

Katherine S. Harris+ and Fredericka Bell-Berti++

Arthur Bronstein, in his book The Pronunciation of American English (1960), follows the convention of dividing the sounds of the language into two classes--the consonants and the vowels. Within this rubric, he assigns the glottal stop [ʔ] and the glottal fricative [h] to the consonant class, as many other authors do. To choose a few examples, [ʔ] is described as a "glottal plosive" and [h] as a "breathed glottal fricative" by Daniel Jones (1956); and [ʔ] as a "laryngeal stop" and [h] as a "laryngeal open consonant" by Heffner (1949). The authors thus make the tacit assumption that these sounds share some property with the stops and fricatives, and contrast, in some manner, with vowels. In part, this view is a consequence of their distributional properties (Andresen, 1968), and, indeed their role in the syllable. However, this decision leaves us with the further problem of deciding what syllables are, within which the consonants and vowels may have roles. To continue with our sampling of phonetics texts, we find Malmberg (1963) and MacKay (1978) observing that, although phoneticians may differ on the definition of a syllable, the untrained speaker of a language usually has a clear idea of the number of syllables in an utterance, and this intuitive reality suggests that there must be some corresponding articulatory reality. For convenience, we will ignore the problems of the more general definitions of the syllable (Pulgram, 1970; Bell & Hooper, 1978), though we note that the problem of finding articulatory meaning for the syllable is made more acute by the failure of efforts to find easy distributional definitions.

Modern physiological research on the syllable begins with the work of R. H. Stetson (1951), who suggested that the syllable was physiologically defined by an initiating and a terminating burst of activity from the muscles of the chest wall, the internal and external intercostal muscles, resulting in a distinct chest pulse for each syllable. This attractive concept was effectively torpedoed by the classic experimental work of Ladefoged and his colleagues (Ladefoged, 1967), who were able to show that there were not discrete bursts of muscle activity corresponding to individual syllables and, indeed, that the manner of interaction of muscular and non-muscular forces in the expiratory cycle made the idea of a syllable based on separate muscular syllable pulses theoretically implausible. More recently, attempts have been made to salvage the concept of an articulatory syllable by assuming that its boundaries may be discovered by careful examination of the activity of the articulators, rather than the respiratory muscles.

---

*To be published in L. Raphael, C. Raphael, & M. Valdovinos (Eds.), Language and cognition: Essays in honor of Arthur J. Bronstein. New York: Plenum Press.

+Also The Graduate School, City University of New York.

++Also St. John's University.

Acknowledgment. This work was supported by NINCDS Grants NS-13617 and NS-13870, and BRS Grant RR-05596 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-70 (1982)]

Many current theories stem from the work of Kozhevnikov and Chistovich (1965), originators of the concept of the articulatory syllable defined by coarticulation. In brief, they suggested that all elements in a single syllable are co-produced. As a consequence, for example, if a syllable contains a rounded vowel, the consonants associated with the syllable would be likely to take on "rounding" attributes. As a correlate, one might suppose that in sequences of an unrounded-vowel syllable followed by a rounded-vowel syllable, an examination of rounding characteristics of the intervocalic consonants might permit the specification of a syllable boundary. In fact, Kozhevnikov and Chistovich suggest an "articulatory" syllable consisting of a vowel and its preceding consonant string. This basic suggestion has been amplified by Gay, who finds that in a VCV string, the articulatory movement toward a second vowel begins at, but never earlier than, the onset of the first intervocalic consonant (Gay, 1978); in other words, the syllable boundary is marked in coarticulatory terms.

Support has been provided for this idea by the so-called "trough phenomenon" (Bell-Berti & Harris, 1974; Gay, 1975). Briefly, it has been shown that if two rounded vowels of the same phonetic specification are produced in sequence, with a single consonant or string of consonants unspecified for rounding between the vowels, as in [utú], the lip muscles will relax between the two vowels, so that the consonants are produced with only partly rounded lips. The same phenomenon can be demonstrated, as well, in sequences like [ipí], where the tongue, which must be raised and fronted for the two identical front vowels, relaxes in association with production of the [p], although the conventional, or feature, description of [p] does not specify a tongue position for the consonant. In both cases, there are two "vowel" gestures, one apparently for each syllable. However, for reasons of economy of production, one might expect a "held" gesture for the second of the two vowels, since the production of the intervening consonantal gesture does not appear to be in conflict with the vowel.

While these facts can be used to argue against some models of coarticulation (Bell-Berti & Harris, 1981), they provide support for coarticulatory marking of syllable boundaries if a trough, indicating a consonant gesture, is formed at all syllable boundaries. In the textbook descriptions of phonetic sequences we provided earlier, we understood that a syllable boundary must occur somewhere in the sequence VCV. The trough phenomenon provides evidence of boundary marking because a vowel-to-vowel gesture, which might, apparently, be produced continuously, is not. If [h] and [ʔ] are consonants, they should interrupt a vowel-to-vowel sequence in the same way that [t] production interrupts vowel rounding.

The general hypothesis is that the "trough" phenomenon is a general syllable boundary marker. We wanted to examine [h] and [ʔ] for the two syllable sequences where the original observations of the trough phenomenon were made. We ask--"Do [h] and [ʔ] cause relaxation of the tongue for [i] sequences" and "Do [h] and [ʔ] cause relaxation of lip protrusion for [u] (or [ɔ]) sequences?"

At present, the most effective way of observing the movements of the tongue is in lateral view cineradiography. We have made extensive observations of tongue movements using a special purpose facility, the x-ray microbeam installation at the University of Tokyo (see Kiritani, Itoh, & Fujimura, 1975).[1] For the purposes of the present discussion, we merely note that the output of the system is a series of plots of the x and y coordinates of the position of pellets affixed to the articulators. The speaker was a male native of southeastern New York State, with no pronounced speech defects.

Figure 1 shows the position of the y coordinate for two pellets as a function of time for three nonsense syllable sequences, [əpihípə], [əpiʔípə] and [əpipípə]. An examination of these three tokens, and others like them that vary in stress and speaking rate, leads to the general impression that a trough is substantially less likely for [ʔ] and [h] than [p]; some samples of [h] show a trough, but most do not. Of course, more quantitative observations are necessary.

It is somewhat easier to observe the movement of the lips in the production of rounded vowels. While it is possible to use x-ray methods, an easier technique is to observe the forward protrusion of the lips in rounded vowel production either by monitoring movies of the lips in profile, or by recording the output of a suitably-placed strain gauge.

Figure 2 shows the lip movement for the sequences [lɔʔɔl] and [lɔtɔl]. Unfortunately, we did not examine the sequence [lɔhɔl]. The speaker was a female native of the Washington, D.C. area, with normal articulation. The recording shows the output of a strain gauge placed on the lower lip in such a way that forward movement of the lip causes bending of the plate (Abbs & Gilbert, 1973).[2] An examination of the figure suggests that there is a trough in the lip-protrusion curve for [t], but not for [ʔ].

Unfortunately, as with many experimental facts, the results just described may be interpreted in several not-mutually-exclusive ways. One possibility is that there is no coarticulatory definition of the syllable boundary. A second possibility is that the "laryngeal" stops [h] and [ʔ], do not form a class with [t] and [p] so that [h] and [ʔ] are not "true" consonants and thus cannot lead to boundaries even if [VhV] and [VʔV] are judged to be disyllabic. A third possibility is that existence of a trough is some sort of a positive articulatory requirement for each phone for which it occurs. Such an approach is taken by Engstrand (1981); he suggests that the lip relaxation associated with [s] and [t] between rounded vowels may arise as a consequence of the aerodynamic prerequisites of these consonant sound types, rather than as a consequence of some general consonant property, or their syllabic position. Presumably, then, by analogy, lip relaxation fails to occur for sequences in which a glottal stop occupies the intervocalic position, because there is no acoustic requirement for such a maneuver. If the argument is accepted, we must then search for those acoustic requirements that specify the details of tongue position for a bilabial stop, in the environment of high front vowels. While it may seem, on the face of it, somewhat unparsimonious to search for two separate acoustic arguments for the appearance of the trough in the two environments, there is no a priori reason to discard the possible explanation.

Observations like those of this experiment substantially restrict the field over which we can apply any "theory" of coarticulation, or of syllabification. Nonetheless, we have ample evidence that the articulatory require-

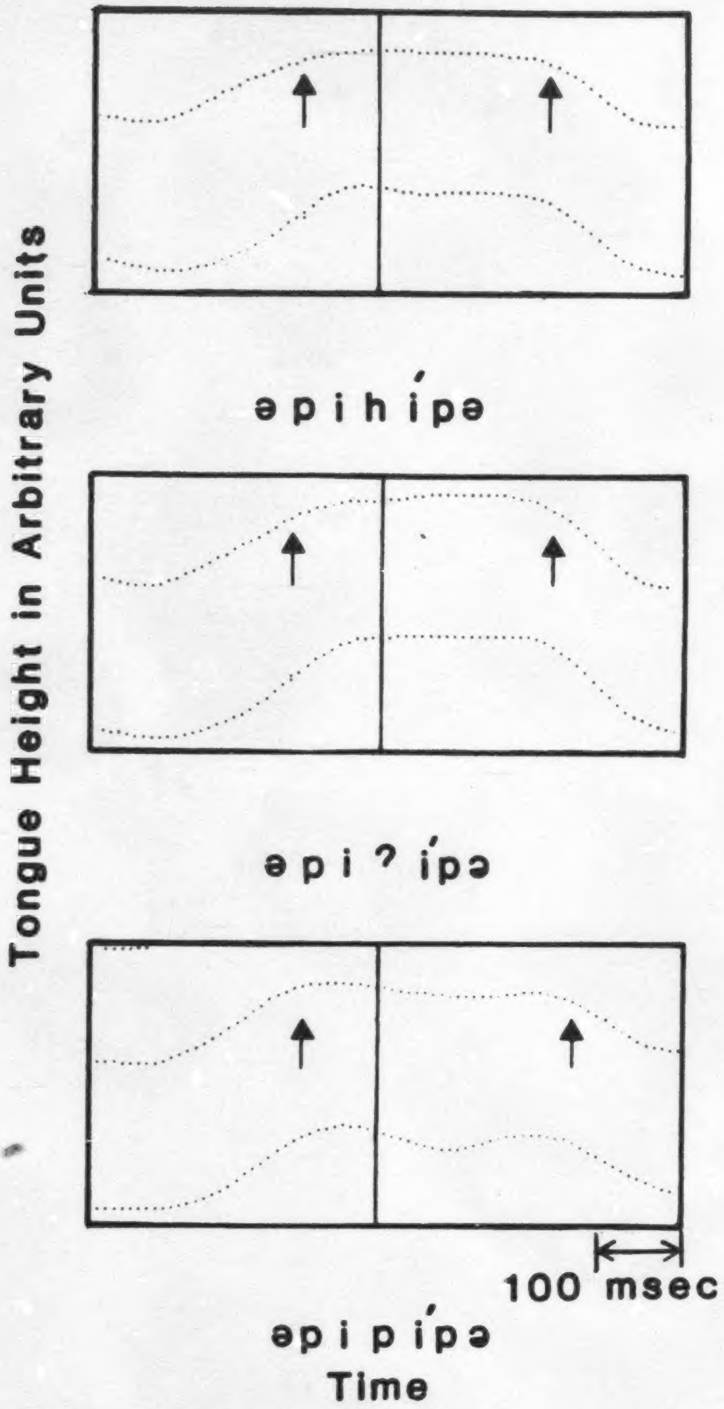


Figure 1. X-ray microbeam traces for the syllables [əpipípə], [əpihípə] and [əpiʔípə]. The plots show the vertical coordinate of a pellet on the tongue blade and mid-tongue position. Coordinate values with larger y values show greater tongue height. The long vertical line on each trace shows the time of the end of voicing for the first [i]. The two upward-pointing arrows show the beginning and end of the two-vowel sequence.

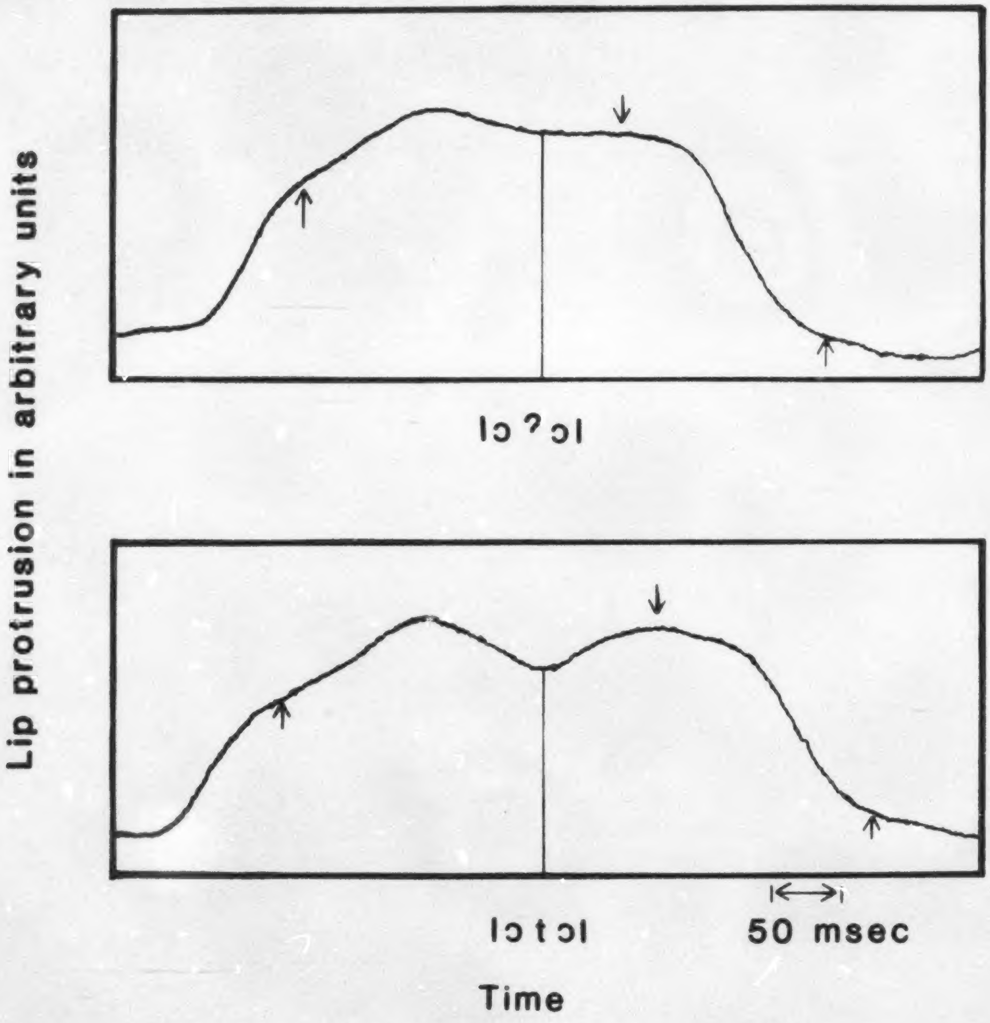


Figure 2. Output of a strain gauge transducer on the lower lip for the syllables [lɔʔɔl] and [lɔtɔl]. The trace shows the forward movement of the lips for rounding during vowel production. Coordinate values increase for greater forward movement of the lip. Line and arrows indicate the same acoustic events as in Figure 1.

ments of a given phone are at least broad enough to allow some contextual variation. It remains for the future, then, for us to develop a theory of syllabification and coarticulation using evidence gathered from the articulatory domain with a net whose mesh has a smaller gauge than that which has produced our present views.

REFERENCES

Abbs, J. H., & Gilbert, B. N. A strain gauge transduction system for lip and jaw motion in two dimensions: Design criteria and calibration data. *Journal of Speech and Hearing Research*, 1973, *16*, 248-256.

Andresen, B. S. *Pre-glottalization in English standard pronunciations*. Oslo, Norway: Norwegian University Press, 1968.

Bell, A., & Hooper, J. B. Issues and evidence in syllabic phonology. In A. Bell & J. B. Hooper (Eds.), *Syllables and segments*. New York: Elsevier North-Holland, 1978.

Bell-Berti, F., & Harris, K. S. More on the motor organization of speech gestures. *Haskins Laboratories Status Report on Speech Research*, 1974, *SR-37/38*, 73-78.

Bell-Berti, F., & Harris, K. S. A temporal model of speech production. *Phonetica*, 1981, *38*, 9-20.

Bronstein, A. J. *The pronunciation of American English*. New York: Appleton-Century Crofts, 1960.

Engstrand, O. Acoustic constraints of invariant input representation? An experimental study of selected articulatory movements and targets. *RUUL 7* (Reports from Uppsala University Department of Linguistics), 1981, 67-94.

Gay, T. Some electromyographic measures of coarticulation in VCV utterances. *Haskins Laboratories Status Report on Speech Research*, 1975, *SR-44*, 137-146.

Gay, T. Articulatory units: Segments or syllables? In A. Bell & J. B. Hooper (Eds.), *Syllables and segments*. Amsterdam: North-Holland, 1978.

Heffner, R. M. S. *General phonetics*. Madison, Wisc.: University of Wisconsin, 1949.

Jones, D. *An outline of English phonetics* (8th ed). New York: Dutton, 1956.

Kiritani, S., Itoh, K., & Fujimura, O. Tongue-pellet tracking by a computer-controlled x-ray microbeam system. *Journal of the Acoustical Society of America*, 1975, *57*, 1516-1520.

Kozhevnikov, R. V., & Chistovich, L. *Speech: Articulation and perception*. (Translated from Russian). Report 30, 543. Washington, D.C.: Joint Publications Research Service, 1965.

Ladefoged, P. *Three areas of experimental phonetics*. London: Oxford University Press, 1967.

MacKay, I. *Introducing practical phonetics*. Boston, Mass.: Little Brown, 1978.

Malmberg, B. *Phonetics*. New York: Dover, 1963.

Pulgram, E. *Syllable, word, nexus, cursus*. The Hague: Mouton, 1970.

Stetson, R. H. *Motor phonetics: A study of speech movements in action* (2nd ed.). Amsterdam: North Holland, 1951.

FOOTNOTES

[1] We are grateful to Dr. Masayuki Sawashima and the staff of the Research Institute of Logopedics and Phoniatrics at the University of Tokyo for their help with these experiments, and to Dr. Osamu Fujimura, Dr. Joan Miller, and Dr. Winston L. Nelson of Bell Laboratories for their help with the preparatory data processing and analysis facilities for the output of the Tokyo System.

[2] We are grateful to Dr. Sandra Hamlet of the University of Maryland for making these recordings at her experimental facility, and to Dr. Maureen Stone for her help in data analysis.

VOWEL INFORMATION IN POSTVOCALIC FRICTIONS*

D. H. Whalen+

Abstract. When the postvocalic frictions of [s] and [š] are excerpted and combined with vocalic segments having inappropriate formant transitions, vowel quality, or both, the fricative percept is determined by the noise. However, there is often a perception of a diphthong in the vowel. This phenomenon was explored for the vowels [a, i, o, u] preceding the fricatives [s] and [š]. In the first of two experiments, all combinations of the vocalic segments and frictions were presented for identification of the vowel. The perception of diphthongs occurred much more often on mismatches of vowel quality than of transition, indicating that there is substantial vowel information in the friction. In the second experiment, just the frictions of the syllables were presented, with subjects trying to identify the missing vowel. The high vowels [i] and [u] were reliably identified, while identifications of [a] and [o] were at chance. This result agrees with previous studies of initial fricatives (Yeni-Komshian & Soli, 1981). Fricative noises from [i] and [u] were responsible for the large majority of diphthong percepts in Experiment 1. These results illustrate that fricative noises contain considerable information about preceding high vowels.

INTRODUCTION

In the production of a phonetic string, both anticipatory and perseverative coarticulation occur. The resulting intermingling of phonetic cues makes the extraction of acoustic segments that are all the cues for one phone and cues only for that phone almost impossible (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Two of the most extractable phones are [s] and [š]. These fricatives are realized by an intense noise that is usually distinct

---

*A version of this paper was presented at the Annual Meeting of the Linguistic Society of America, December, 1981, New York, New York.

+Also Yale University.

Acknowledgments. I would like to thank Alvin M. Liberman, Leigh Lisker, and Bruno Repp for help in writing this paper. The research and writing of this paper were supported by NICHD Grant HD-01994.

[HASKINS LABORATORIES: Status Report on Speech Research SR-70 (1982)]

from the accompanying segments, and this noise is quite identifiable as to the fricative produced (Harris, 1958; Heinz & Stevens, 1961; Hughes & Halle, 1956; Yeni-Komshian & Soli, 1981). Yet there is also a substantial and perceivable residue of vowel information (LaRiviere, Winitz, & Herriman, 1975; Yeni-Komshian & Soli, 1981). In addition, there is fricative information that remains in the vocalic segment (Mann & Repp, 1980; Whalen, 1981).

Although the vowel information in these intial fricatives leads to correct identifications of some vowels from the friction alone, it is not highly salient. Not only are the percentages for correct identification of the vowel well below those for identification of the fricative, this vowel information also does not override the information contained in the vocalic segment when the two cues are made to conflict. Indeed, such mismatches seldom result in any directly perceivable effect. Whalen (in press) explores subtler effects of such mismatches that show up only in reaction time paradigms.

The present work examines the corresponding effects of coarticulation in vowel-fricative syllables. Pilot observations suggested that cross-spliced syllables in which vowel quality cues in the frictions and in the vowel itself conflict often give rise to a diphthong percept. Experiment 1 examines this in detail for the vowels [a, i, o, u] and the fricatives [s] and [š]. The second experiment assesses the identifiability of the preceding vowel from the friction alone, complementing earlier work on initial fricatives.

EXPERIMENT 1

Procedure

Materials. A male native speaker of English recorded ten tokens of each of the syllables [as], [aš], [is], [iš], [os], [oš], [us], and [uš] on magnetic tape. Lip configuration was maintained into the frication. The rounded vowels were not intentionally diphthongized. The stimuli were low-pass filtered at 10 kHz and digitized at a sampling rate of 20 kHz. Two tokens of each syllable were chosen so that both the vocalic portion and the friction would be of equal duration in all eight. A vocalic segment duration of 200 msec was found naturally in eight syllables. Seven were shortened by cutting off between 10 and 50 msec from the first part of the vowel; the resulting abrupt onset did not sound unnatural. The eighth modified vocalic portion was lengthened 20 msec by repeating its first pitch pulse three times. The frictions were 250 msec in duration; nine were shortened by removing between 10 and 50 msec from near the end of the signal.

Once the tokens had been selected and the durations equalized, each friction was combined with each vocalic segment, including the original. This gave four main categories for the 256 stimuli: 1) The vowel was the same as the one the friction was originally produced with (henceforth, "the vowel matched the original vowel" or just "the vowel was matched") and the vocalic formant transitions were appropriate to the fricative ("the transitions were matched"); 2) the vowel was matched but the transitions were mismatched; 3) the vowel was mismatched and the transitions were matched; and 4) both vowel and transition were mismatched.

The stimuli were randomized and recorded on magnetic tape for presentation. The interstimulus interval was 3.5 seconds, with 6 seconds after every ten stimuli.

Subjects. Ten subjects were run. Seven were researchers at Haskins Laboratories who were phonetically trained and/or had extensive experience in speech research. The other three were native speakers of English who had volunteered for experiments at Haskins Laboratories, and were paid for their participation.

Apparatus and procedure. Subjects heard the stimuli over TDH-39 headphones. They recorded their identifications of the vowel on the answer sheet as follows: Non-diphthongized vowels were simply written as "a," "i," "o," or "u," with the phonetic value of each being explained to the naive subjects. Diphthongized vowels were written as a sequence of two of these symbols, whether or not they characterized the exact nature of the offglide.

Results

Each subject gave four judgments for each combination of vowel and original vowel (of the friction). The number of diphthongs perceived by each subject ranged from two to sixty (out of 256 judgments). Misidentifications of the main vowel were excluded from the analysis; they comprised 2.9% of the data.

All four of the vowel categories were given as the second vowel (or offglide). The number of times a particular vowel was identified as the offglide is given in Table 1. There were few reports of [a] and [o] offglides, so these were excluded from the statistical analysis.

Results obtained with initial fricatives would lead us to expect that a mismatch of transition would give rise to diphthong percepts. With some tokens of initial fricatives, joining [š] transitions to a friction from [s] results in the perception of a [y] glide. In the current stimuli, there were eighty syllables in which the vowel quality was matched but the transitions were mismatched. In only one of these cases (the vocalic segment of [oš] with the friction of [os]) was a diphthong perceived. With these stimuli, then, the transitions were not the cause of the diphthong percepts.

Of the 204 diphthongs analyzed, 74.5% occurred when the original vowel and the offglide percept were both [i] or both [u]. If we include those cases where the vowel with which the fricative was produced agreed in rounding with the offglide (i.e., [a] giving an [i] offglide and [o] giving an [u] offglide), 93.6% of the cases are accounted for. Thus a large proportion of the responses showed agreement in rounding between the vowel and the original vowel.

Table 1

Number of Offglide Percepts, by Fricative Category, for the Four Vowel Qualities

	Fricative was		Total
	s	š	
# of [a] offglides	1	3	4
# of [o] offglides	3	6	9
# of [u] offglides	33	15	48
# of [i] offglides	137	19	156

Discussion

It is clear that the vowel quality information in the friction is primarily responsible for the diphthong that is perceived. There was one "oi" judgment (mentioned above) when the transition was inappropriate, but overall, mismatch of transition did not seem to be a contributing factor.

That the offglides were overwhelmingly judged as [i] and [u] is no surprise. These are not only the common offglides of American English, but they are also articulatorily the easiest offglides to make in a brief time. (Remember that subjects were to classify offglides that approached [i] and [u] as [i] and [u] rather than being more exact.) To get an [a] percept, for example, there must be tongue and jaw lowering. When there is a fricative to follow, this gesture requires much more time to accomplish than an offglide to, say, [i], since [i] is close to the semi-closed position that [s] or [š] will require. For this reason, listeners rarely reported [a] offglides in the present stimuli.

EXPERIMENT 2

The preponderance of [i] and [u] offglide percepts in Experiment 1 was explained in erms of articulatory constraints on offglides. However, it may be that these vowels leave more of a coarticulatory trace in the final frictions than do [a] and [o]. If the frictions contain information only about the high vowels, it would not be surprising that high offglides are perceived. This hypothesis is tested in Experiment 2.

Procedure

Materials. The frictions of Experiment 1 were isolated, and 16 repetitions of each were randomized and recorded on magnetic tape. The interstimulus interval was 3500 msec.

Subjects. The ten subjects of Experiment 1 participated.

Apparatus and procedure. Subjects heard the stimuli over TDH-39 headphones. They indicated which vowel must have preceded the fricative by depressing one of four buttons, labeled "a," "i," "o," or "u." The phonetic value of each symbol was explained to the naive subjects. The buttons were connected to a computer, which provided immediate feedback for correct responses.

Results

Overall, the vowel was correctly identified 41.25% of the time. This was significantly above chance ($t(9) = 4.09$, $p < .005$). Of the four vowels, however, only [i] and [u] were identified at above chance levels (see Table 2); this was true with both [s] and [š] (Table 2).

The four vowels can be compared on the features of rounding and (relative) height. Subjects identified the roundness of the missing vowel correctly significantly more often than chance (see Table 3; $\chi^2 = 322.04$, $p < .001$). Subjects also did better than chance on the height feature (Table 3; $\chi^2 = 48.354$, $p < .001$). It appeared that rounding was correctly identified more often than height. A sign test for the ten subjects shows this difference to be significant (9 of 10, $p = .011$).

The two features behaved differently with the different fricatives. When the fricative was [s], more unrounded vowel judgments were given, while [š] elicited more rounded judgments (Table 4; $\chi^2 = 322.04$). Similarly, the vowel judged to have preceded an [s] was judged as high and [š] as low more often than chance would dictate (Table 4; $\chi^2 = 48.354$).

Discussion

The identifiability of the vowels from the frictions agrees well with previous work. The addition of [o] to the previously studied [a], [i], and [u] allows us to make some tentative comparisons along the features of rounding and height. These comparisons indicate that rounding is more easily reconstructed from these frictions than height. This is presumably the perceptual reflection of the acoustic shaping imposed on the friction by the rounded lips. A relatively lower noise would lead the listener to think that the missing vowel must have been rounded. While the present data tend to bear this out, the higher proportion of round vowel responses to [š] noises confuses the issue. Since the [š] noise is lower in frequency than that of [s], the comparison of relative height within [s] or within [š] noises becomes more difficult. A study that presented only [s] noises or only [š] noises would test the presumed salience of rounding more directly.

Table 2

Test for Above Chance Identification of Individual Vowels and Vowel-Fricative Combinations, Experiment 2.

	% friction	% correct correct (both [s] and [s])
[a]s	34.48	34.22
[a]š	33.96	
[i]s	75.62*	61.82*
[i]š	47.96*	
[o]s	21.38	28.57
[o]š	35.84	
[u]s	35.63*	40.31*
[u]š	45.00*	

*Significantly better than chance ($p < .01$).

Table 3

Number of Judgments of Round or Unround, High or Low Vowels.

		Fricative was produced after a vowel that was:				
		unround	round		low	high
vowel identified was	unround	903	373	low	641	469
	round	451	826	high	633	810

Table 4

Number of Judgments of Round or Unround, High or Low Vowels.

		Fricative produced was:				
		s	š		s	š
vowel identified was	unround	841	513	low	471	639
	round	437	762	high	807	636

The listeners' responses were not based on direct perception of a vowel in the friction but rather on educated guesses. The only vowels that may have been directly perceived were the [i]'s from [s]. Many subjects reported hearing those as a whispered vowel followed by a fricative. Thus the information in these frictions, though demonstrably present, is not strong enough to build a solid percept in isolation.

GENERAL DISCUSSION

The two experiments described here combine to show that there is vowel information in the noise portion of final fricatives that is sufficient to give actual vowel (offglide) percepts when the fricative noise is preceded by a mismatched vowel. Considering Experiment 1 by itself, we postulated phonotactic and articulatory reasons for the preponderance of [i] and [u] offglides in the diphthong percepts. Taking Experiment 2 into account, we can see that these are the two vowels that are inherently more identifiable from the friction. Thus the vowels that leave the strongest coarticulatory trace, as measured by identifiability in Experiment 2, are the most common diphthong percepts in Experiment 1. In addition, those frictions that prompted the most correct identification of the missing vowel were the frictions that gave rise to the majority of the diphthong percepts (156 of 204, as noted above).

The two major effects seen in the present experiments, that high vowels and rounded vowels coarticulate the most with final [s] and [š], are clearly based in the possibilities of articulation. Since the narrow constriction necessary for producing [i] and [u] is close to that needed for the fricatives, the two gestures can affect each other more easily than with [o] and [a]. Since the lips are not primary articulators for [s] or [š], they can maintain their rounding through the fricative uninterruptedly. Although [u] is both high and round and [i] only high, [i] was recognized more frequently. This is due to two factors: First, rounding seems to be detectable both in its presence and it absence so that lack of rounding is as much of a cue as its presence. Second, the constriction for [u], though near the roof of the mouth, is not as near to the final point of articulation of the fricatives. The constriction for [i], on the other hand, is quite near that of [s]. This seems to allow the articulators to maintain their position, rather than having to break it off (as with [as]). The result is high identifications for [is] (75.62%).

The greater identifiability of the high vowels is apparent in the perception of diphthongs in syllables with mismatched cues as well. While the diphthongs of English usually end in a high vowel (thus providing a possible bias in the perception), they may do so for articulatory reasons. In an offglide, we expect less than full vowel quality; yet if there is a consonant following, we must also have a quick movement into the articulation appropriate for it. The high vowels allow this movement much more easily than the low. This, combined with the greater coarticulation discovered for high vowels in the vowel identification test, accounts for the preponderance of [i] and [u] offglides in the diphthong percepts.

Together, these results show that it is inappropriate to call the vocalic segment of a syllable the vowel (cf. Repp, 1981). Just as there is consonant information in the vocalic segment (for fricatives, see Mann & Repp 1980; Whalen, 1981), so there is vowel information in the friction of final fricatives. Therefore, not only is the vocalic segment not entirely a vowel, it is not the entire vowel either. While the vowel information in the friction is not sufficient to override information in vocalic segments, Experiment 1 shows us that it can, in the proper circumstances, be perceived as vowel information. Only further experimentation will tell whether it is powerful enough to affect ambiguous vocalic segments, thus demonstrating its cue value in a more traditional manner.

REFERENCES

Harris, K. S. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1958, *1*, 1-7.

Heinz, J. M., & Stevens, K. N. On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America*, 1961, *33*, 589-596.

Hughes, G. W., & Halle, M. Spectral properties of fricative consonants. *Journal of the Acoustical Society of America*, 1956, *28*, 303-310.

LaRiviere, C. J., Winitz, H., & Herriman, E. The distribution of perceptual cues in English prevocalic fricatives. *Journal of Speech and Hearing Research*, 1975, *18*, 613-622.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, *74*, 431-461.

Mann, V. A., & Repp, B. H. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, 1980, *28*, 213-228.

Repp, B. H. On levels of description in speech research. *Journal of the Acoustical Society of America*, 1981, *69*, 1462-1464.

Whalen, D. H. Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. *Journal of the Acoustical Society of America*, 1981, *69*, 275-282.

Whalen, D. H. Subcategorical phonetic mismatches slow phonetic judgments. *Haskins Laboratories Status Report on Speech Research*, in press.

Yeni-Komshian, G. H., & Soli, S. D. Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, 1981, *70*, 966-975.

II. PUBLICATIONS

III. APPENDIX

PUBLICATIONS

Alfonso, P. J., & Baer, T. Dynamics of vowel articulation. *Language and Speech*, in press.

Crowder, R. G. A common basis for auditory sensory storage in perception and immediate memory. *Perception & Psychophysics*, 1982, *31*, 477-484.

Davis, W., & Kelso, J. A. S. An analysis of "invariant characteristics" in the motor control of normal and Down's Syndrome subjects. *Journal of Motor Behavior*, 1982, *14*.

Harris, K. S., & Bell-Berti, F. On consonants and syllable boundaries. In L. J. Raphael, C. Raphael, & M. Valdovinos (Eds.), *Language and cognition.: Essays in honor of Arthur J. Bronstein*. New York: Plenum, in press.

Kelso, J. A. S., Holt, K. G., Rubin, P., & Kugler, P. N. Patterns of human interlimb coordination emerge from the properties of nonlinear, limit cycle oscillatory processes: Theory and data. *Journal of Motor Behavior*, 1981, *13*, 226-261.

Kelso, J. A. S., & Tuller, B. H. "Compensatory articulation" under conditions of reduced afferent information: A dynamic formulation. *Journal of Speech and Hearing Research*, in press.

Kelso, J. A. S., & Tuller, B. H. A dynamical basis for action systems. In M. S. Gazzaniga (Ed.), *Handbook of cognitive neuroscience*. New York: Plenum, in press.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. On the control and coordination of naturally developing systems. In J. A. S. Kelso & J. E. Clark (Eds.), *The development of movement control and coordination*. Chichester: John Wiley & Sons, 1982.

Metz, D., Whitehead, R., & McGarr, N. S. Physiological aspects of speech produced by deaf persons. *Audiology*, 1982, *7*(3), 35-48.

Osberger, M. J., & McGarr, N. S. Speech production characteristics of the hearing impaired. In N. Lass (Ed.), *Speech and language: Advances in basic research and practice*. New York: Academic Press, 1982.

Repp, B. H., & Mann, V. A. Fricative-stop coarticulation: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 1982, *71*, 1562-1567.

Tuller, B., Harris, K. S., & Kelso, J. A. S. Stress and rate: Differential transformations of articulation. *Journal of the Acoustical Society of America*, 1982, *71*, 1534-1543.

Tuller, B., Kelso, J. A. S., & Harris, K. S. Interarticulator phasing as an index of temporal regularity in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 1982, *8*, 460-472.

Yumoto, E., Gould, W. J., & Baer, T. Harmonics-to-noise ratio as an index of the degree of hoarseness. *Journal of the Acoustical Society of America*, 1982, *71*, 1544-1550.

APPENDIX

DTIC (Defense Technical Information Center) and ERIC (Educational Resources Information Center) numbers:

Status Report		DTIC	ERIC
SR-21/22	January - June 1970	AD 719382	ED 044-679
SR-23	July - September 1970	AD 723586	ED 052-654
SR-24	October - December 1970	AD 727616	ED 052-653
SR-25/26	January - June 1971	AD 730013	ED 056-560
SR-27	July - September 1971	AD 749339	ED 071-533
SR-28	October - December 1971	AD 742140	ED 061-837
SR-29/30	January - June 1972	AD 750001	ED 071-484
SR-31/32	July - December 1972	AD 757954	ED 077-285
SR-33	January - March 1973	AD 762373	ED 081-263
SR-34	April - June 1973	AD 766178	ED 081-295
SR-35/36	July - December 1973	AD 774799	ED 094-444
SR-37/38	January - June 1974	AD 783548	ED 094-445
SR-39/40	July - December 1974	AD A007342	ED 102-633
SR-41	January - March 1975	AD A013325	ED 109-722
SR-42/43	April - September 1975	AD A018369	ED 117-770
SR-44	October - December 1975	AD A023059	ED 119-273
SR-45/46	January - June 1976	AD A026196	ED 123-678
SR-47	July - September 1976	AD A031789	ED 128-870
SR-48	October - December 1976	AD A036735	ED 135-028
SR-49	January - March 1977	AD A041460	ED 141-864
SR-50	April - June 1977	AD A044820	ED 144-138
SR-51/52	July - December 1977	AD A049215	ED 147-892
SR-53	January - March 1978	AD A055853	ED 155-760
SR-54	April - June 1978	AD A067070	ED 161-096
SR-55/56	July - December 1978	AD A065575	ED 166-757
SR-57	January - March 1979	AD A083179	ED 170-823
SR-58	April - June 1979	AD A077663	ED 178-967
SR-59/60	July - December 1979	AD A082034	ED 181-525
SR-61	January - March 1980	AD A085320	ED 185-636
SR-62	April - June 1980	AD A095062	ED 196-099
SR-63/64	July - December 1980	AD A095860	ED 197-416
SR-65	January - March 1981	AD A099958	ED 201-022
SR-66	April - June 1981	AD A105090	ED 206-038
SR-67/68	July - December 1981	**	ED 212-010
SR-69	January - March 1982	**	ED 214-226

Information on ordering any of these issues may be found on the following page.

**DTIC and/or ERIC order numbers not yet assigned.

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International
Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Language Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
Haskins Laboratories
270 Crown Street
New Haven, CT 06510

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
N/A

3. REPORT TITLE
Haskins Laboratories Status Report on Speech Research, SR-70, April-June, 1982

4. DESCRIPTIVE NOTES *(Type of report and, inclusive dates)*
Interim Scientific Report

5. AUTHOR(S) *(First name, middle initial, last name)*
Staff of Haskins Laboratories, Alvin M. Liberman, P.I.

6. REPORT DATE
June, 1982

7a. TOTAL NO. OF PAGES
328

7b. NO. OF REFS
616

8a. CONTRACT OR GRANT NO.
HD-01994 NS13870
HD-16591 NS13617
N01-HD-1-2420 NS18010
RR-05596
PRF-8006144
BNS-8111470

9a. ORIGINATOR'S REPORT NUMBER(S)
SR-70 (1982)

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
None

10. DISTRIBUTION STATEMENT
Distribution of this document is unlimited*

11. SUPPLEMENTARY NOTES
N/A

12. SPONSORING MILITARY ACTIVITY
See No. 8

13. ABSTRACT

This report (1 April-30 June) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

--Exploring the functional significance of physiological tremor: A biospectroscopic approach
--Differences between experienced and inexperienced listeners to deaf speech
--A language-oriented view of reading and its disabilities
--Phonetic factors in letter detection: A reevaluation
--Categorical perception: Issues, methods, findings
--Short-term recall by deaf signers of American Sign Language: Implications of encoding strategy for order recall
--A common basis for auditory sensory storage in perception and immediate memory
--Phonological awareness and verbal short-term memory: Can they presage early reading problems?
--Initiation versus execution time during manual and oral counting by stutterers
--Trading relations in the perception of speech by five-year-old children
--The role of the strap muscles in pitch lowering
--Phonetic validation of distinctive features: A test case in French
--On consonants and syllable boundaries
--Vowel information in postvocalic frictions

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

UNCLASSIFIED
Security Classification

A-31408

*This document contains no information not freely available to the general public. It is distributed primarily for library use.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT

Speech Perception:
  deaf speech, experience
  categorical perception, review
  auditory memory, short-term
  children, trading relations
  features, phonetic, French
  vowels, fricatives, postvocalic

Speech Articulation:
  stuttering, initiation, execution, time
  strap muscles, pitch
  consonants, syllables

Reading:
  disability, language
  letters, detection, phonetic factors
  American Sign Language, memory, short-term
  phonological awareness

Motor Control:
  tremor, function, spectrum